多层递阶方法及其应用

韩志刚 编著

科学出版社

多层递阶方法及其应用

韩志刚 编著

科学出版社

内容简介

多层递阶预报方法是本书作者从80年代初开始研究并应用的一种预报方法．由于这种方法充分考虑了被预报系统的时变特性，所以较大地提高了预报精度．几年来，这种方法已成功地应用于石油、农业、天气、经济系统的预报以及工业过程自适应控制并收到了较好的效果．

本书较系统地介绍了多层递阶预报方法及其应用，叙述了一些与其有关的基础知识，如系统辨识的基本理论、多层时间序列分析和多层建模方法等等．

本书可作为自动控制、经济系统预报、天气预报等部门的实际工作者以及有关专业的大学生和研究生的参考书．

多层递阶方法及其应用

韩志刚 编著

责任编辑 李淑兰 杨 艳

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1989年1月第一版 开本：850×1168 1/32

1989年1月第一次印刷 印张：7 3/4

印数：0001—4,150 字数：199,000

ISBN 7-03-000675-5/TP·44

定价：4.90元

序　言

在当今这个时代，人们正变得越来越关心未来，因而对预测技术的研究广泛地引起了人们的重视.由于计算机技术的飞速发展，使得定量预测有了强有力的计算工具，因而对于定量预测的理论和方法的研究也有了很大进展．但是目前常用的各种定量预测方法，都或多或少地存在着一定的缺点，其主要原因是，用一个非时变的预报模型来预报一个时变动态系统的状态或输出，会引起较大的预报误差，而且随着预报步长的增加，这种误差增加得很快．为此，我们从 1982 年展开了多层递阶预报方法及其应用的研究．

建立预报模型是定量预报的一个关键性步骤．把以控制系统设计为目的而建立的控制模型进行适当改变，就可以作为以预报为目的的预报模型．但预报问题有着自己的特殊性，如果忽略了这种特殊性，用处理控制问题的某些观点和方法来处理预报问题，将可能使预报效果受到一定的影响.

控制系统，特别是具有反馈环节的控制系统，要求其数学模型对可能出现的系统状态变化有较广泛的适应性，即要求模型具有"相空间"的某种大范围的特征．然而预报问题的解决，基本上是依据系统状态及与之有关的某些量的过去和现在的一组观测值来推断它的未来．此时，我们仅涉及到系统的一条"轨线"．这就是说，预报模型不一定要求具有"相空间"的大范围特征．但从另一方面看，预报问题总是要涉及一定的"时间尺度"，特别是"长期"预报，其"时间尺度"就更大了．所以说，"相空间"中的局部性和"时间尺度"的大范围性，是预报问题所具有的特殊性．因而用预报模型进行辨识和预测，都应该有相应的特殊方法，这就是我们提出的多层建模方法的思想基础.

多层递阶预报方法已在石油、农业、天气、经济和交通等方面

得到了应用，并收到了较好的效果．目前，多层递阶方法的理论和应用仍处于发展阶段．

本书对多层建模方法、多层递阶预报方法及其应用情况进行了介绍，作为基础也介绍了系统辨识的基本理论与已有的重要预报方法．我之所以有勇气写这本书，是因为受到了自动控制界许多同行们的鼓励．

张志方教授仔细地阅读了原稿并提出了宝贵意见．陈振宇高级工程师和王恩平副研究员对于本书的写作给予了大力支持，赵永胜、张恩恕和王杰等同志在他们各自的工作中成功地应用了多层递阶方法，汤兵勇和郭一新同志帮助我完成了大量的计算工作，王洪桥、吴敏惠等同志帮助我整理了书稿，在此向他们表示谢意．

为了便于实际工作者阅读，对本书中的某些定理没有进行严格的推导，对此有兴趣的读者可参阅有关文献．由于作者水平有限，错误在所难免，希望得到批评指正．

编著者

1987年3月1日

目 录

第一章 基本数学模型……1

§1.1 随机序列……1

§1.2 平稳随机序列……4

§1.3 滑动平均序列……7

§1.4 ARMA 模型……11

§1.5 多维随机序列和多维 ARMA 模型……14

§1.6 ARMA 模型的推广……17

§1.7 状态空间模型……20

第二章 非时变参数系统的辨识……23

§2.1 最小二乘法……23

§2.2 最小二乘估计的递推算法……26

§2.3 辅助变量法……31

§2.4 广义最小二乘法和推广的最小二乘法……35

§2.5 预报误差估值与极大似然法……39

§2.6 随机逼近算法……46

§2.7 常见的递推算法所对应的微分方程……52

§2.8 常见的递推算法的收敛性……57

§2.9 模型定阶的 AIC 准则……62

§2.10 模型定阶的 FPE 准则……66

第三章 解决预报问题的古典方法……70

§3.1 时间序列的预报问题……70

§3.2 瓦尔德分解……74

§3.3 平稳时间序列预报的维纳-柯尔莫哥洛夫方法……79

§3.4 时间序列预报的博克斯-詹金斯方法……86

§3.5 具有 MA 噪声的系统预报的奥斯特隆姆方法……89

§3.6 自校正预报方法……92

§3.7 卡尔曼预报方法……………………………………………… 95
第四章 动态系统时变参数的辨识………………………… 100
§4.1 预报模型的参数辨识准则…………………………………… 100
§4.2 预报模型的参数估计算法…………………………………… 103
§4.3 多输出系统的参数估计……………………………………… 107
§4.4 多输出系统的分解…………………………………………… 112
§4.5 参数估计算法的改进………………………………………… 116
§4.6 参数估计算法的分析………………………………………… 119
§4.7 预报模型参数的完全辨识…………………………………… 125
§4.8 典型形式 $y_k = \varphi(k)^\tau A f(k) + w(k)$ …………………… 126
§4.9 时变参数的辨识与典型形式………………………………… 129
§4.10 典型形式的结构确定 ……………………………………… 132
§4.11 算法的递推化问题 ………………………………………… 134
第五章 多层递阶建模与预报………………………………… 136
§5.1 时间序列的多层分析方法…………………………………… 136
§5.2 多层递阶预报方法…………………………………………… 139
§5.3 高层时间序列建模与预报的具体方法……………………… 142
§5.4 预报结果的误差分析………………………………………… 146
§5.5 相关噪声干扰下的多层递阶预报方法……………………… 149
§5.6 预报问题的特殊性…………………………………………… 154
§5.7 一致小误差预报……………………………………………… 156
§5.8 预报模型参数估值的初值选择……………………………… 159
§5.9 非线性模型线性化的一种途径……………………………… 163
第六章 天气与农业系统的预报……………………………… 167
§6.1 天气系统及其预报问题……………………………………… 167
§6.2 关于春季平均气温的长期预报……………………………… 169
§6.3 西太平洋夏季各月份副高面积指数长期预报……………… 172
§6.4 西太平洋夏季各月份副高西伸脊点长期预报……………… 181
§6.5 农作物单位面积产量的预报模型…………………………… 190
§6.6 试验预报的结果与分析……………………………………… 195
§6.7 主要农作物单位面积产量向后18年的预报结果 ………… 197
§6.8 小麦单位面积产量的多层递阶自回归预报………………… 201
§6.9 关于农业总产值的预报……………………………………… 203

第七章　工业经济系统的预报…………………………… 207

§7.1　油田产量的预报………………………………………… 207

§7.2　产品销售量的预报……………………………………… 214

§7.3　工业总产值的预报……………………………………… 218

第八章　多层递阶方法在自适应控制中的应用……………… 223

§8.1　参数预报自适应控制及其在农业中的应用…………… 223

§8.2　稳定的参数自适应控制………………………………… 226

§8.3　含未知部分的自适应控制模型的补偿方法…………… 230

§8.4　补偿自适应最小方差控制……………………………… 232

§8.5　多层递阶参数自适应控制在造纸机上的应用………… 234

第一章　基本数学模型

§1.1　随 机 序 列

预报(或预测)的基本问题就是要依据事物的发展变化规律推断它的未来．不掌握事物的发展变化规律，就没有办法对它进行准确的预报．可以说，认真研究和正确掌握事物发展变化规律，是预报工作的基础．

为正确地揭示和掌握事物的发展变化规律，必须对它进行一系列的观测或实验，从中提取必要的信息．观测或实验的结果往往被表示成一系列数据

$$y_1,\ y_2,\ \cdots,\ y_N$$

从概率论和数理统计学的观点来看，这一系列数据是描写所考察事物某种品性之随机序列的一个实现的组成部分．

所谓随机序列，是指一族随机变量

$$\{x_k,\ k\in T\}$$

此处，$T=\{\cdots,\ -2,\ -1,\ 0,\ 1,\ 2,\ \cdots\}$ 或者 $T=\{0,\ 1,\ 2,\ \cdots\}$．事实上，一个随机序列是定义在 $\Omega\times T$ 上的二元函数

$$x(\omega,\ k)\quad \omega\in\Omega;\ k\in T$$

而 Ω 是概率空间 $(\Omega,\ \mathscr{F},\ P)$ 的基本事件集（基本空间），T 称为随机序列的参量集．对 $x(\omega,\ k)$ 进行一次观测，相当于固定一个 $\omega\in\Omega$．从而得到它的一个实现

$$y_1,\ y_2,\ \cdots,\ y_n,\ \cdots$$

随机序列 $\{x_k,\ k\in T\}$ 的整数变量 k，往往表示某种时间，例如第 k 小时、第 k 天、第 k 年、第 k 次等等，所以人们常常把随机序列称为时间序列．

类似地，如果参量集是一个连续系统，则相应的随机变量族

$\{x_t, t\in T\}$，称为随机过程.

由于随机序列或随机过程是由一族随机变量构成的，因此它有着自己的统计特征.

1. 随机序列的分布

设 $\{x_k\}$ 是一个随机序列，对于参量值的任何一个有限的集合 $\{t_1, t_2, \cdots, t_n\}\in T$，相应的随机变量组成了一个随机向量

$$(y_{t_1}, y_{t_2}, \cdots, y_{t_n})^\tau$$

τ 表示转置. 它的联合分布函数

$$F_n(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)$$
$$= P[x_{t_1}\leqslant y_1, x_{t_2}\leqslant y_2, \cdots, x_{t_n}\leqslant y_n]$$

称为随机序列的一个有限维分布函数.

由于随机序列中含有可数无穷多个随机变量，因此它的有限维分布函数的数目是无限的，并形成一个族

$$\{F_n(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)\}$$

我们称此为随机序列的有限维分布函数族.

随机序列的有限维分布函数应该满足下述条件:

1) 对称性: 对于 1, 2, …, n 的任何一种重新排列 r_1, r_2, $\cdots$, r_n, 皆有

$$F_n(y_{r_1}, y_{r_2}, \cdots, y_{r_n}; t_{r_1}, t_{r_2}, \cdots, t_{r_n})$$
$$= F_n(y_1, y_2, \cdots, y_n; t_1, t_2, \cdots, t_n)$$

2)相容性: 对于任意的 $m<n$, 我们有

$$F_m(y_1, y_2, \cdots, y_m; t_1, t_2, \cdots, t_m)$$
$$= \lim_{\substack{y_{m+1}\to+\infty \\ y_n\to+\infty}} F(y_1, y_2, \cdots, y_m, y_{m+1}, \cdots, y_n; t_1, t_2, \cdots, t_n)$$

这就是说，对于任何随机序列，必有与之相应的有限维分布函数族. 反过来可以证明，如果给定了一族有限维分布函数，且它满足对称性和相容性条件，则必能找到一个随机序列，且它与有限维

分布函数族相对应.

有限维分布函数族完全描述了随机序列的分布特征，但是全部掌握这一族分布函数通常是十分困难的.从应用的观点来看,有时也没有这种必要. 而仅需掌握它的某些参数特征，就可以描述随机序列的主要统计特性了. 因此，以下我们来介绍随机序列的参数特征.

2. 均值序列(函数)

对于每个 $k\in T$, 随机变量 x_k 的均值如果存在,则

$$\mu(k)=E\{x_k\}$$

形成一个序列 $\{\mu(k)\}$，于是称 $\{\mu(k)\}$ 为随机序列 $\{x_k\}$ 的均值序列.

3. 协方差函数

对于任何的两个时刻 $k, l\in T$, 如果

$$E\{(x_k-\mu(k))(x_l-\mu(l))\}=\text{cov}\{x_k, x_l\}$$

存在,则称其为随机序列 $\{x_k\}$ 的协方差函数，记作 $r(k, l)$. 特别地称 $\text{cov}\{x_k, x_k\}$ 为随机序列 $\{x_k\}$ 的方差函数. 在 $\{x_k\}$ 取复值的情形，我们有

$$r(k, l)=E\{(x_k-\mu(k))\overline{(x_l-\mu(l))}\}$$

显然有

$$|r(k, l)|\leqslant r(k, k)^{\frac{1}{2}}r(l, l)^{\frac{1}{2}} \tag{1.1}$$

4. 相关函数

随机序列的相关函数 $\rho(k, l)$ 定义为

$$\rho(k, l)=\frac{r(k, l)}{r(k, k)^{\frac{1}{2}}\cdot r(l, l)^{\frac{1}{2}}} \tag{1.2}$$

协方差函数和相关函数都是随机序列 $\{x_k\}$在不同时刻 k, l 的随机变量 x_k, x_l 相关程度的数量特征.

§1.2 平稳随机序列

平稳随机序列在时间序列分析中占有重要地位．它描写了这样一类现象：其统计特性不随时间的推移而变化．随机序列的平稳性可分为狭义的和广义的两种．

1. 狭义平稳序列

设 $\{x_k, k=0, \pm 1, \pm 2, \cdots\}$ 是一个随机序列，如果对任何的 $k_1, k_2, \cdots, k_n$ 和整数 h，随机向量

$$(x_{k_1}, x_{k_2}, \cdots, x_{k_n})$$

与随机向量

$$(x_{k_1+h}, x_{k_2+h}, \cdots, x_{k_n+h})$$

皆有相同的分布，就说该序列是一个狭义的平稳序列．

不难看出，如果狭义平稳序列的均值函数和协方差函数都存在，则必有

1) $\mu(k)=\mu(l)=\mu$，对一切的 k 和 l；

2) $r(k, l)=r(k-l, 0)$，对一切的 k 和 l．

这就是说，具有前二阶矩的狭义平稳序列，其均值函数为常数，协方差函数仅依赖于时刻 k 与 l 之差．于是，如果令 $\tau=k-l$，则有 $r(k, l)=r(\tau, 0)$．下面将 $r(\tau, 0)$ 简记为 $r(\tau)$．

2. 广义平稳序列

设 $\{x_k, k\in T\}$ 是一个具有二阶矩的随机序列，如果它满足

1) 对一切 k，皆有 $\mu(k)=c$，c 是与 k 无关的常数；

2) 对一切整数 k 和 l，皆有

$$r(k, l)=r(k-l, 0)=r(k-l)$$

则说该随机序列是一个广义平稳序列．

显然，具有二阶矩的狭义平稳序列必为广义平稳序列．但前二阶矩一般并不能决定随机变量的整个分布，所以广义平稳序列

一般并不是狭义平稳的．然而，对于正态的情形，则例外．

我们说一个随机序列是正态的（高斯的），如果它的一切有限维分布都是正态的．因为正态分布完全被它的前二阶矩所决定，所以正态的广义平稳过程亦必是狭义平稳的．

3. 协方差函数和谱函数

对于广义平稳序列而言，协方差函数 $r(k)$ 是它的重要数字特征．不难验证，协方差函数有如下性质：

1） $r(0) \geqslant 0$；

2） $r(-k) = r(k)$；

3） $|r(k)| \leqslant r(0)$；

4）对任意正整数 n，实数 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 及整数 $k_1, k_2, \cdots, k_n$，皆有

$$\sum_{i=1}^{n} \sum_{j=1}^{n} r(k_i - k_j)\alpha_i\alpha_j \geqslant 0$$

或者 n 阶方阵

$$\begin{pmatrix} r(0) & r(k_1 - k_2) & \cdots & r(k_1 - k_n) \\ r(k_2 - k_1) & r(0) & \cdots & r(k_2 - k_n) \\ \vdots & \vdots & & \vdots \\ r(k_n - k_1) & r(k_n - k_2) & \cdots & r(0) \end{pmatrix}$$

是非负定的．

反之可以证明，具有上述所有性质的实数序列$\{r(k)\}$，必为某平稳序列的协方差函数．我们称具有上述所有性质的序列$\{r(k)\}$为非负定序列．

定理 1.1 $\{r(k)\}$ 为非负定序列的充要条件是它能够表示为

$$r(k) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i k\lambda} dF(\lambda) \tag{1.3}$$

其中 $F(\lambda)$ 是 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 上的一个单调不减函数，在适当的正则化情况下，$F(\lambda)$ 由 $r(k)$ 唯一确定．如果有 $f(\lambda)$ 使 $dF(\lambda) =$

$f(\lambda)d\lambda$，则 (1.3) 式可写为

$$r(k)=\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}f(\lambda)d\lambda \tag{1.4}$$

这个定理的证明可参看文献 [1]. 我们称 $F(\lambda)$ 为相应的平稳序列的谱分布函数或简称为谱函数. 如果 $f(\lambda)$ 存在,则称它为相应的平稳序列的谱密度函数或简称谱密度.

进一步还可以证明:

定理 1.2 如果协方差函数 $\{r(k)\}$ 满足

$$\sum_{k=-\infty}^{+\infty}|r(k)|<+\infty$$

则谱密度函数一定存在，且

$$f(\lambda)=\sum_{k=-\infty}^{+\infty} r(k)e^{-2\pi ik\lambda} \tag{1.5}$$

4. 平稳序列的谱展式

设 $\{x_t, t\in T\}$ 是一个随机过程. 如果对于满足 $s_1<t_1\leqslant s_2<t_2$ 的参量值 s_1, t_1, s_2, t_2, 恒有

$$E\{(x_{t_1}-x_{s_1})(x_{t_2}-x_{s_2})\}=0$$

则称 $\{x_{t_1}, t\in T\}$ 是一个具有直交增量的过程. 可以证明如下的定理:

定理 1.3 每个广义平稳序列 $\{x_k, k=0, \pm1, \pm2, \cdots\}$ 皆具有谱展式

$$x_k=\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}dZ(\lambda) \tag{1.6}$$

此处 $\left\{Z(\lambda), -\frac{1}{2}\leqslant\lambda\leqslant\frac{1}{2}\right\}$ 是一个具有直交增量的过程，并且满足

$$E\{|dZ(\lambda)|^2\}=dF(\lambda)$$

$F(\lambda)$ 是该平稳随机序列的谱函数.在适当正则化的条件下,$Z(\lambda)$ 由 $\{x_k\}$ 唯一确定. 详细的讨论可参看文献[1].

§ 1.3 滑动平均序列

1. 平稳序列的线性运算

设 $\{x_k\}$ 是一个零均值的平稳序列，$a_1, a_2, \cdots, a_n$ 是常数，则

$$y_k = \sum_{l=1}^{n} a_l x_{k-l} \qquad \forall k$$

称为由 $\{x_k\}$ 经线性运算所得到的序列. 可以证明它仍是一个零均值的平稳序列. 一般地，如果实数列 $\{a_k\}$ 满足

$$\sum_{k=-\infty}^{+\infty} |a_k| < \infty \tag{1.7}$$

则

$$y_k = \sum_{l=-\infty}^{+\infty} a_l x_{k-l}$$

所构成的序列 $\{y_k\}$ 仍为零均值的平稳序列.

事实上，由于有条件 (1.7) 式，所以

$$E\{y_k\} = \sum_{l=-\infty}^{+\infty} a_l E\{x_{k-l}\} = 0$$

进一步，由于

$$E\{y_k^2\} = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} a_i a_j r_x(i-j)$$

$$\leqslant r_x(0) \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} |a_i| \cdot |a_j|$$

$$= r_x(0)\left(\sum_{l=-\infty}^{+\infty} |a_l|\right)^2 < +\infty \qquad \forall k$$

故 $\{y_k\}$ 具有有限的二阶矩，其中 $r_x(k)$ 表示 $\{x_k\}$ 的协方差函数. 又由于

$$E\{y_k \cdot y_l\} = E\left\{\sum_{i=-\infty}^{+\infty} a_i x_{k-i} \cdot \sum_{j=-\infty}^{+\infty} a_l x_{l-j}\right\}$$

$$= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} a_i a_j E\{x_{k-i} \cdot x_{l-j}\}$$

$$= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} a_i a_j r_x(k-l-i+j)$$

即它仅依赖于 $k-l$，故 $\{y_k\}$ 是一个平稳序列.

2. 滑动平均

设 $\{e_k\}$ 是一个零均值的随机序列，并且满足

$$\operatorname{cov}\{e_k, e_l\} = \sigma^2 \cdot \delta_{k,l}$$

其中 $\delta_{k,l}$ 满足

$$\delta_{k,l} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

则称 $\{e_k\}$ 为零均值的白噪声. 显然它也是一个平稳序列. 谱密度函数 $f_e(\lambda)$ 必存在，并且有

$$f_e(\lambda) = \sum_{k=-\infty}^{+\infty} \sigma^2 \delta(k) e^{-2\pi i k \lambda}$$

其中 $\delta(k)$ 满足

$$\delta(k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

所以有

$$f_e(\lambda) = \sigma^2$$

这就是说，白噪声序列的谱密度是常数.

如果

$$y_k = \sum_{l=-\infty}^{+\infty} a_l e_{k-l} \tag{1.8}$$

其中 $\{e_k\}$ 是零均值的白噪声，$E\{e_k^2\} = \sigma^2$. 而 $\{a_l\}$ 满足

$$\sum_{l=-\infty}^{+\infty} |a_l| < +\infty$$

则说 $\{y_k\}$ 是 $\{e_k\}$ 的滑动平均(或滑动和)，$\{y_k\}$ 是一个平稳随机

序列．较重要的是下述的单边滑动平均序列

$$y_k = \sum_{l=0}^{\infty} a_l e_{k-l} \tag{1.9}$$

其中 $\{a_k\}$ 满足 $\sum_{l=0}^{\infty} |a_l| < +\infty$．不难算出，它的协方差函数 $r_y(k-l)$ 为

$$\begin{aligned} r_y(k-l) &= E\left(\sum_{i=0}^{\infty} a_i e_{k-i} \sum_{j=0}^{\infty} a_j e_{l-j}\right) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i a_j \mathrm{E}\{e_{k-i} e_{l-j}\} \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i a_j \delta_{k-i,l-j} \sigma^2 \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i a_j \delta(k-l-i+j) \sigma^2 \\ &= \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+(k-l)} \end{aligned}$$

相应的谱密度函数为

$$\begin{aligned} f_y(\lambda) &= \sum_{k=-\infty}^{+\infty} r_y(k) e^{-2\pi i k \lambda} \\ &= \sigma^2 \sum_{k=-\infty}^{+\infty} \sum_{l=0}^{\infty} a_l a_{l+k} e^{-2\pi i k \lambda} \end{aligned}$$

当 $k<0$ 时，只要令 $a_k = 0$，则 (1.9) 式也可以写成双边滑动平均的形式．于是有

$$\begin{aligned} f_y(\lambda) &= \sigma^2 \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} a_l a_{l+k} e^{-2\pi i k \lambda} \\ &= \sigma^2 \sum_{l=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} a_l a_{l+k} e^{-2\pi i k \lambda} \\ &= \sigma^2 \sum_{l=0}^{+\infty} \sum_{k=-l}^{+\infty} a_l a_{l+k} e^{-2\pi i k \lambda} \end{aligned}$$

$$= \sigma^2 \sum_{l=0}^{+\infty} \sum_{k=-l}^{+\infty} a_l \overline{e^{-2\pi i k\lambda}} a_{l+k} e^{-2\pi i(l+k)\lambda}$$

$$= \sigma^2 \left(\sum_{l=0}^{+\infty} a_l \overline{e^{-2\pi i l\lambda}} \right) \left(\sum_{k=-l}^{+\infty} a_{l+k} e^{-2\pi i(k+l)\lambda} \right)$$

$$= \sigma^2 \left| \sum_{l=0}^{\infty} a_l e^{-2\pi i l\lambda} \right|^2$$

如果置

$$C(z) = \sum_{l=0}^{\infty} a_l z^l$$

则有

$$f_y(\lambda) = \sigma^2 |C(e^{-2\pi i\lambda})|^2$$

特别地,对于下述的有限和

$$y_k = \sum_{l=0}^{m} a_l e_{k-l} \tag{1.10}$$

我们有

$$f_y(\lambda) = \sigma^2 |C(e^{-2\pi i\lambda})|^2$$

其中

$$C(z) = \sum_{l=0}^{m} a_l z^l$$

可见,此时 $f_y(\lambda)$ 是关于 $e^{-2\pi i\lambda}$ 的多项式.

3. 具有有理谱密度的序列

设 $\{x_k\}$ 是一个平稳序列,如果其谱密度函数存在,而且具有形式

$$f(\lambda) = \sigma^2 \left| \frac{C(e^{-2\pi i\lambda})}{A(e^{-2\pi i\lambda})} \right|^2$$

其中 $\sigma^2 > 0$, 而 $A(z)$, $C(z)$ 皆为 z 的实系数多项式. 它们之间无公因子,并且 $A(z)$, $C(z)$ 的零点皆在单位圆外。我们就称 $\{x_k\}$ 为具有有理谱密度的平稳序列。

序列 (1.10) 是具有有理谱密度的平稳序列，在那里 $A(z)=1$.

注意到谱密度的一般形式

$$f(\lambda)=\sum_{k=-\infty}^{+\infty} r(k)e^{-2\pi ik\lambda}$$

显然它是 $e^{-2\pi i\lambda}$ 的某个函数. 在满足一定条件时，$f(\lambda)$ 可以具有连续性. 然而我们知道，任何连续函数皆可用有理函数逼近，所以具有有理谱密度的序列是有一定的广泛意义的.

§1.4 ARMA 模型

用随机序列或过程，仅能对客观事物本身(例如预报对象，预报因子等)的一些数量变化特征进行描述，而要深入研究其发展变化规律，一般来说，应该建立起它所满足的某种数学模型. 下面所介绍的 ARMA 模型，是描述很广泛的一类平稳序列的变化规律的较好模型.

设 $\{x_k\}$ 是一个随机序列，$\{e_k\}$ 是零均值的白噪声序列，如果 $\{x_k\}$ 满足.

$$x_k+a_1x_{k-1}+\cdots+a_nx_{k-n}=e_k+c_1e_{k-1}+\cdots+c_me_{k-m} \tag{1.11}$$

而多项式

$$A(z)=1+a_1z+\cdots+a_nz^n$$
$$C(z)=1+c_1z+\cdots+c_mz^m$$

满足条件

1) $A(z)$ 与 $C(z)$ 互质;

2) $a_n\neq 0$, $c_m\neq 0$;

3) $A(z)$ 和 $C(z)$ 的零点都在单位圆外.

此处还假定对一切 $l>0$, x_{k-l} 与 e_k 不相关，则序列 $\{x_k\}$ 称为 n 阶自回归-m 阶滑动平均混合序列，模型(1.11) 称为 n 阶自回归-m 阶滑动平均混合模型，简记为 ARMA(n,m) 模型. n 称为

自回归阶数，m 称为滑动平均阶数．实参数 $a_1, a_2, \cdots, a_n$ 称为自回归系数，$c_1, c_2, \cdots, c_m$ 称为滑动平均系数．

特别地，如果 $n=0$，则模型称为纯滑动平均模型，记作 MA(m)；如果 $m=0$，则模型称为自回归模型，记作 AR(n)．类似地，我们称满足模型 AR(n) 的序列为 AR(n) 序列．满足 MA(m) 的序列为 MA(m) 序列．

如果引入一步延迟算子 q^{-1}

$$q^{-1}x_k = x_{k-1},\quad q^{-1}e_k = e_{k-1}$$

则 ARMA(n, m) 模型 (1.11) 可简记为

$$A(q^{-1})x_k = C(q^{-1})e_k \tag{1.11$'$}$$

其中

$$A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + \cdots + a_nq^{-n}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + c_2q^{-2} + \cdots + c_mq^{-m}$$

以下的定理说明了 ARMA(n, m) 模型与具有有理谱密度的平稳序列之间的关系．

定理 1.4 设 $A(e^{-2\pi\lambda}) \neq 0$．如果 $\{x_k\}$ 是随机差分方程

$$A(q^{-1})x_k = C(q^{-1})e_k$$

的平稳解，则 $\{x_k\}$ 有如下形式的谱密度

$$f(\lambda) = \sigma^2\left|\frac{C(e^{-2\pi i\lambda})}{A(e^{-2\pi i\lambda})}\right|^2 \tag{1.12}$$

其中 $\sigma^2 = E\{e_k^2\}$．反之如果均值为零的平稳序列 $\{x_k\}$ 的谱密度具有 (1.12) 式的形式，则它必满足随机差分方程

$$A(q^{-1})x_k = C(q^{-1})e_k$$

证明 注意到 $\{e_k\}$ 和 $\{x_k\}$ 的平稳性，我们有

$$e_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}dz_e(\lambda)$$

其中 $z_e(\lambda)$ 为相应于 e_k 的具有直交增量的过程，它满足

$$E\{|dz_e(\lambda)|^2\} = \sigma^2 d\lambda$$

$$x_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}dz_x(\lambda)$$

$z_x(\lambda)$ 为相应于 x_k 的具有直交增量的过程，它满足

$$E\{|dz_x(\lambda)|^2\} = dF(\lambda)$$

$F(\lambda)$ 为 $\{x_k\}$ 的谱函数. 把上述的谱展式代入随机差分方程

$$A(q^{-1})x_k = C(q^{-1})e_k$$

中,我们得

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}A(e^{-2\pi i\lambda})dz_x(\lambda) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}C(e^{-2\pi i\lambda})dz_e(\lambda)$$

由于上式对一切 $k = 0, \pm 1, \pm 2, \cdots$ 皆成立,所以有

$$|A(e^{-2\pi i\lambda})|^2 dF(\lambda) = |C(e^{-2\pi i\lambda})|^2\sigma^2 d\lambda$$

由此得出

$$f(\lambda) = \frac{dF(\lambda)}{d\lambda} = \sigma^2\left|\frac{C(e^{-2\pi i\lambda})}{A(e^{-2\pi i\lambda})}\right|^2$$

反过来,如果零均值的平稳序列 $\{x_k\}$ 具有 (1.12) 式形式的谱密度,则它的协方差函数必有形式

$$r(k) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}f(\lambda)d\lambda = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}\left|\frac{C(e^{-2\pi i\lambda})}{A(e^{-2\pi i\lambda})}\right|^2\sigma^2 d\lambda \tag{1.13}$$

设 $z_e(\lambda)$ 是 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 上的具有直交增量的过程, 它满足 $E\{|dz_e(\lambda)|^2\} = \sigma^2 d\lambda$. 用 $z_x(\lambda)$表示平稳序列 $\{x_k\}$ 的谱展式所对应的具有直交增量的过程,则由 (1.13) 式我们有

$$E\{|dz_x(\lambda)|^2\} = dF(\lambda) = f(\lambda)d\lambda = \sigma^2\left|\frac{C(e^{-2\pi i\lambda})}{A(e^{-2\pi i\lambda})}\right|^2 d\lambda = E\left\{\left|\frac{C(e^{-2\pi i\lambda})}{A(e^{-2\pi i\lambda})}dz_e(\lambda)\right|^2\right\}$$

所以有

$$x_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}\frac{C(e^{-2\pi i\lambda})}{A(e^{-2\pi i\lambda})}dz_e(\lambda)$$

置

$$e_k=\int_{-\frac{1}{2}}^{\frac{1}{2}}e^{2\pi ik\lambda}dz_e(\lambda)$$

因为

$$E\{e_k\bar{e}_l\}=\sigma^2\int_{-\frac{1}{2}}^{\frac{1}{2}}e^{2\pi i(k-\lambda)}d\lambda=\begin{cases}\sigma^2 & k=l\\0 & k\neq l\end{cases}$$

故 $\{e_k\}$ 是白噪声序列．不难直接验证，x_k 满足随机差分方程

$$\begin{aligned}&x_k+a_1x_{k-1}+\cdots+a_nx_{k-n}\\&=e_k+c_1e_{k-1}+\cdots+c_me_{k-m}\end{aligned}$$ ■

因为随机差分方程 (1.11)的平稳解和具有有理谱密度的序列间有一一对应的关系，所以用 ARMA (n, m) 序列可以很好地拟合具有连续谱的零均值平稳序列．这说明了 ARMA(n, m) 模型具有一定的广泛性．

§ 1.5 多维随机序列和多维 ARMA 模型

多维随机序列和一维的情形一样，有着明显的实际和理论意义．

依赖于时间变量 k 的 p 维随机向量

$$\mathbf{x}(k)=\{x_1(k),\ x_2(k),\ \cdots,\ x_p(k)\}^\tau$$

即称为一个 p 维随机序列，$k\in T$，T 是参数集，$T=\{0, 1, 2, \cdots\}$ 或者 $T=\{0, \pm1, \pm2, \cdots\}$．

p 维的随机序列可以看成是由 p 个互相关联的一维随机序列 $\{x_i(k)\}$ $i=1, 2, \cdots, p$ 组合而成的．

1. 多维随机序列的数量特征

和一维的情形一样，多维随机序列 $\{\mathbf{x}(k)\}$ 也有表征它的统计特性的重要数量特征，它们是均值函数和协方差矩阵．

p 维随机序列 $\{\mathbf{x}(k)\}$ 的均值函数是一个向量值函数

$$E\{\mathbf{x}(k)\}=[E\{x_1(k)\},\ E\{x_2(k)\},\cdots,E\{x_p(k)\}]^\tau=\boldsymbol{\mu}(k)$$

显然 $\{\boldsymbol{\mu}(k)\}$ 是一个普通的 p 维向量序列.

p 维随机序列 $\{\mathbf{x}(k)\}$ 的协方差矩阵定义如下：设它的每个分量皆有有限的二阶矩，则称矩阵

$$E\{[\mathbf{x}(k)-\boldsymbol{\mu}(k)]\overline{[\mathbf{x}(l)-\boldsymbol{\mu}(l)]}^{\tau}\}$$

为 $\{\mathbf{x}(k)\}$ 的协方差矩阵，记作 $R(k, l)$.

协方差矩阵不仅描述了在时刻 k, l 随机向量 $\mathbf{x}(k)$ 和 $\mathbf{x}(l)$ 之间的相关程度，而且也描述了 $\mathbf{x}(k)$ 的各分量之间的关联关系. 特别地，我们称

$$R(k, k)=E\{(\mathbf{x}(k)-\mu(k))\overline{(\mathbf{x}(k)-\mu(k))}^{\tau}\}$$

为 p 维随机序列的方差矩阵.

2. 多维平稳序列

这里我们只考虑广义情形. 设 $\{\mathbf{x}(k)\}$ 是一个多维的随机序列. 如果它的每个分量的二阶矩皆存在，而且满足条件

1) 对任意 k 皆有

$$E\{\mathbf{x}(k)\}=\boldsymbol{\mu} \quad \boldsymbol{\mu}\text{为常值 } p \text{ 维向量}$$

2) 对于任意的 k 和 l, 皆有

$$E\{(\mathbf{x}(k)-\boldsymbol{\mu})\overline{(\mathbf{x}(l)-\boldsymbol{\mu})}^{\tau}\}=R(k-l, 0)$$

则说 $\{\mathbf{x}(k)\}$ 是一个 p 维平稳序列. 此时把它的协方差函数 $R(k-l, 0)$ 简记为 $R(k-l)$.

我们说两个随机序列 $\{\mathbf{x}_1(k)\}$ 和 $\{\mathbf{x}_2(k)\}$ 是平稳相关的，如果对任何整数 k, l, h 皆有

$$\operatorname{cov}\{\mathbf{x}_1(k+h), \mathbf{x}_2(l+h)\}=\operatorname{cov}\{\mathbf{x}_1(k), \mathbf{x}_2(l)\}$$

显然，p 维平稳序列的每个分量皆为一维平稳序列，而且任何两个分量序列之间是平稳相关的.

设 $\{\mathbf{e}(k)\}$ 是一个 p 维的零均值的平稳序列，如果其协方差函数 $R(k)$ 具有形式

$$R(k)=D\delta(k)$$

其中 D 为正定阵，则说 $\{\mathbf{e}(k)\}$ 是一个零均值的白噪声.

3. 多维的 ARMA 模型

设 $\{\mathbf{x}(k)\}$ 是一个 p 维的随机序列，$\{\mathbf{e}(k)\}$ 是 p 维的零均值白噪声，如果 $\{\mathbf{x}(k)\}$ 满足

$$\mathbf{x}(k) + A_1\mathbf{x}(k-1) + \cdots + A_n\mathbf{x}(k-n) = \mathbf{e}(k) + C_1\mathbf{e}(k-1) + \cdots + C_m\mathbf{e}(k-m) \quad (1.14)$$

其中 $A_1, A_2, \cdots, A_n$　$C_1, C_2, \cdots, C_m$ 皆为 $p\times p$ 的方阵，$A_n \neq 0$，$C_m \neq 0$. 而矩阵多项式

$$\alpha(Z) = I + A_1Z + \cdots + A_nZ^n \quad I \text{ 是 } p\times p \text{ 单位阵}$$

和

$$\gamma(Z) = I + C_1Z + \cdots + C_mZ^m$$

满足 $|\alpha(Z)|$ 与 $|\gamma(Z)|$ 的零点皆在单位圆外，此处 $|\alpha(Z)| = \det \alpha(Z)$，$|\gamma(Z)| = \det\gamma(Z)$. 则说 $\{\mathbf{x}(k)\}$ 是一个 p 维的 ARMA(n, m) 序列. 模型 (1.14) 称为 p 维的 ARMA(n, m) 模型. n 是自回归阶数，m 是滑动平均阶数.

应用算子 q^{-1} 的矩阵多项式 $\alpha(q^{-1})$ 和 $\gamma(q^{-1})$，可以把 (1.14) 改写成形式

$$\alpha(q^{-1})\mathbf{x}(k) = \gamma(q^{-1})\mathbf{e}(k) \quad (1.15)$$

特别地，我们称

$$\mathbf{x}(k) + A_1\mathbf{x}(k-1) + \cdots + A_n\mathbf{x}(k-n) = \mathbf{e}(k)$$

为 AR(n) 模型.

$$\mathbf{x}(k) = \mathbf{e}(k) + C_1\mathbf{e}(k-1) + \cdots + C_m\mathbf{e}(k-m)$$

为 MA(m) 模型.

其实，p 维的 AR, MA 和 ARMA 模型是如下的一般的 p 维单边无限滑动平均模型

$$\mathbf{x}(k) = \sum_{l=0}^{\infty} E_l\mathbf{e}(k-l) \quad (1.16)$$

的特殊情形，此处 $E_0, E_1, \cdots, E_n, \cdots$ 皆为 $p\times p$ 的矩阵. 如果置

$$E(q^{-1}) = \sum_{l=0}^{\infty} E_l q^{-k}$$

则 (1.16) 式可以写成

$$\mathbf{x}(k) = E(q^{-1})\mathbf{e}(k)$$

而对于 AR(n) 模型有：$E(q^{-1}) = [\alpha(q^{-1})]^{-1}$；对于 $MA(m)$ 模型有：$E(q^{-1}) = \gamma(q^{-1})$；对于 ARMA($n$, m) 模型有：$E(q^{-1}) = [\alpha(q^{-1})]^{-1}\gamma(q^{-1})$.

§1.6 ARMA 模型的推广

在预报和控制问题中，常常要涉及一个动态系统．被预报或被控制的随机序列往往作为动态系统的输出，这种输出序列有时表示成 $\{y_k\}$．例如对于 ARMA(n, m) 模型的情形，我们可以把它写成

$$A(q^{-1})y_k = C(q^{-1})e_k$$

其中 $\{e_k\}$ 是零均值的白噪声．上述这个模型描写了所涉及的动态系统的变化规律，y_k 就是这个系统的输出，e_k 是这个系统的随机干扰式的输入，通常 e_k 是不可测量的，$C(q^{-1})e_k$ 是这个系统的随机干扰．在考虑预报与控制问题时，除涉及到动态系统的输出和随机干扰外，还要涉及到可调节的输入和不可调节但可观测的输入，我们把这种输入序列记成为 $\{u_k\}$，通常 u_k 是向量．总之，为了较好地解决预报和控制问题，一般应该考虑具有输出、输入和随机干扰的动态系统的模型．而把无输入仅有输出和随机干扰的动态系统模型作为特殊情形．为此，我们把 AR(n)，MA(m) 和 ARMA(n, m) 模型加以推广．不妨假定 u_k 是一维的．

CAR(n) 模型

$$A(q^{-1})y_k = B(q^{-1})u_k + e_k \tag{1.17}$$

其中 $\{y_k\}$ 是输出序列，$\{u_k\}$ 是输入序列，$\{e_k\}$ 是零均值的白噪声序列．而

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n}$$
$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_m q^{-m}$$

假定多项式 $A(Z)$ 和 $B(Z)$ 没有公因子，$A(Z)$ 的零点在单位圆外。我们称模型 (1.17) 式为带有输入变量的 AR(n) 模型，简称为 CAR(n) 模型。

CMA(m) 模型

$$Y_k = B(q^{-1})u_k + C(q^{-1})e_k \tag{1.18}$$

其中

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_p q^{-p}$$

多项式 $C(Z)$ 的零点皆在单位圆外。模型 (1.18) 式称为具有输入变量的 MA(m) 模型,简称为 CMA(m) 模型。

CARMA(n, m) 模型

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k \tag{1.19}$$

假定多项式 $A(Z)$, $B(Z)$, $C(Z)$ 之间没有公因子，$A(Z)$ 和 $C(Z)$ 的零点皆在单位圆外。我们称模型 (1.19) 式为具有输入变量的 ARMA(n, m) 模型,简称为 CARMA(n, m)模型或 ARMAX (n, m) 模型。

较一般地，我们可以考虑如下的具有 ARMA 噪声的动态调节模型：

$$A(q^{-1})y_k = B(q^{-1})u_k + \frac{C(q^{-1})}{D(q^{-1})}e_k \tag{1.20}$$

其中 $A(Z)$, $B(Z)$ 称为系统多项式,它们是

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n}$$
$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_m q^{-m}$$

而 $C(Z)$ 和 $D(Z)$ 称为噪声多项式,它们是

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_p q^{-p}$$
$$D(q^{-1}) = 1 + d_1 q^{-1} + \cdots + d_l q^{-l}$$

假定多项式 $A(Z)$, $B(Z)$, $C(Z)$, $D(Z)$之间没有公因子,$A(Z)$, $C(Z)$ 和 $D(Z)$ 的零点皆在单位圆外，我们称模型 (1.20) 式为具有 ARMA 噪声的动态调节模型。事实上,如果置

$$\xi_k = \frac{C(q^{-1})}{D(q^{-1})} e_k$$

则有

$$D(q^{-1})\xi_k = C(q^{-1})e_k$$

这是一个 ARMA(l, p) 模型，从而 $\{\xi_k\}$ 是一个 ARMA(l, p) 平稳序列。

模型 (1.19) 式可以看成为模型 (1.20) 式在 $D(Z) = 1$ 时的特殊情形，所以它可以称为具有滑动平均 (MA) 噪声的动态调节模型。与此对应，也可以考虑特殊情形：$C(q^{-1}) = 1$，此时模型 (1.20) 式变为

$$A(q^{-1})y_k = B(q^{-1})u_k + \frac{1}{D(q^{-1})} e_k \tag{1.21}$$

我们把这个模型称为具有自回归 (AR) 噪声的动态调节模型。

更一般地，可以引入如下的转移函数模型：

$$y_k = \frac{B(q^{-1})}{A(q^{-1})} u_k + \frac{C(q^{-1})}{D(q^{-1})} e_k \tag{1.22}$$

其中多项式 $A(q^{-1})$，$B(q^{-1})$，$C(q^{-1})$，$D(q^{-1})$ 满足前面所给出的条件。

上述这些模型，都可以被推广到多维的情形。以转移函数模型为例，我们有

$$\mathbf{y}_k = G(q^{-1})\mathbf{u}(k) + H(q^{-1})\mathbf{e}(k) \tag{1.23}$$

其中 $\mathbf{y}_k$ 为 n 维的输出，$\mathbf{u}(k)$ 是 m 维的输入，$\mathbf{e}(k)$ 是 r 维的零均值的白噪声。$G(Z)$ 和 $H(Z)$ 是 Z 的矩阵值有理函数，$H(Z)$ 满足 $H(0) = I$ (I 是单位阵)，并且 $\det H(Z)$ 的零点皆在单位圆外。

最后，我们来介绍较一般的预报误差模型，先引入新息序列的概念。

定义 设 $\{\mathscr{F}_k\}$ 是概率空间 $(\Omega, \mathscr{F}, P)$ 中的 σ-代数的不减序列。$\{\varepsilon(k)\}$ 是一个随机序列，$\varepsilon(k)$ 关于 $\mathscr{F}_k$-可测。如果有

$$E\{\varepsilon(k)|\mathscr{F}_{k-1}\}=0 \qquad \text{对一切 } k$$
则说 $\{\varepsilon(k)\}$ 是一个关于 $\{\mathscr{F}_k\}$ 的新息序列.

例如：零均值的白噪声序列 $\{e_k\}$ 关于 σ- 代数列 $\mathscr{F}_k=\mathscr{B}(e_1, e_2, \cdots, e_k)$(即$\mathscr{F}_k$ 是由$\{e_1, e_2, \cdots, e_k\}$所导出的$\sigma$-代数)构成一个新息序列.

设 y_k 是输出，u_k 是输入. $\{\varepsilon(k)\}$ 是新息序列. 置

$$\mathbf{Y}_{k-1}=\{y_{k-1}, y_{k-2}, \cdots\}$$

$$\mathbf{U}_k=\{u_k, u_{k-1}, \cdots\}$$

$f[\mathbf{Y}_{k-1}, \mathbf{U}_k, k]$ 是 $\mathbf{Y}_{k-1}$, $\mathbf{U}_k$ 的一个函数，如果有

$$y_k=f[\mathbf{Y}_{k-1}, \mathbf{U}_k, k]+\varepsilon(k) \tag{1.24}$$

并且

$$E\{\varepsilon(k)|\mathbf{Y}_{k-1}, \mathbf{U}_k\}=0$$

则说模型 (1.24) 式是一个向前一步的预报误差模型.

显然，模型 (1.17) 式是一种特殊的预报误差模型.

§ 1.7 状态空间模型

离散时间的随机动态系统的状态方程往往可以写成下述随机差分方程形式

$$\mathbf{x}_{k+1}=f[\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_{k+1}, k] \tag{1.25}$$

其中 $\mathbf{x}_k$ 是 n 维的状态向量，$\mathbf{u}_k$ 是 m 维的控制向量，$\mathbf{w}_k$ 是 p 维的随机噪声，$f[\cdot]$ 是 n 维的向量值函数，一般假定 f 关于它的自变量是连续可微的.

如果 $\mathbf{w}_k$ 是白噪声，并且与初始条件 $\mathbf{x}_0$ 独立，则当 $\mathbf{x}_k$ 已给定时，由方程 (1.25) 看出，$\mathbf{x}_{k+1}$ 仅仅依赖于 $\mathbf{w}_{k+1}$，而 $\mathbf{w}_{k+1}$ 与 $\mathbf{x}_{k-1}, \cdots, \mathbf{x}_0$ 是独立的，所以 $\mathbf{x}_{k+1}$ 在 $\mathbf{x}_k$ 给定时的条件分布与 $\mathbf{x}_{k-1}, \cdots, \mathbf{x}_0$ 无关. 这就是说方程 (1.25) 的解 $\{\mathbf{x}_k\}$ 具有马尔可夫(Markov)性.

如果不存在随机干扰，则方程 (1.25)就是普通的差分方程. 方程(1.25)的一个重要的特殊情形是下列的线性随机差分方程

$$\mathbf{x}_{k+1} = A(k)\mathbf{x}_k + B(k)\mathbf{u}_k + \Gamma(k)\mathbf{w}_{k+1} \tag{1.26}$$

其中 $A(k)$, $B(k)$ 和 $\Gamma(k)$ 分别是 $n \times n$, $n \times m$ 和 $n \times p$ 矩阵, $\mathbf{w}_k$ 是 p 维的高斯白噪声,并且满足

$$E\{\mathbf{w}_k\} = 0 \quad E\{\mathbf{w}_k \mathbf{w}_l^T\} = Q(k)\delta_{k,l}$$

其中 $Q(k)$ 是非负定的矩阵,而

$$\delta_{k,l} = \begin{cases} 0 & k \neq l \\ 1 & k = l \end{cases}$$

如果 x_0 是高斯的,而 $\{\mathbf{u}_k\}$ 或者是非随机的,或者是高斯的,则方程 (1.26) 的解 $\{x_k\}$ 形成一个高斯马尔可夫序列.

进一步,如果 $\{\mathbf{u}_k\}$ 是非随机的,不难看出 $\mathbf{x}_k$ 的均值和协方差阵

$$\bar{\mathbf{x}}_k = E\{\mathbf{x}_k\}$$
$$P(k) = E\{(\mathbf{x}_k - \bar{\mathbf{x}}_k)(\mathbf{x}_k - \bar{\mathbf{x}}_k)^\tau\}$$

满足下述的关系式

$$\bar{\mathbf{x}}_{k+1} = A(k)\bar{\mathbf{x}}_k + B(k)\mathbf{u}_k \tag{1.27}$$

$$P(k+1) = A(k)P(k)A(k)^\tau + \Gamma(k)Q(k+1)\Gamma(k)^\tau \tag{1.28}$$

事实上 (1.26)式两端取均值,注意到

$$E\{\mathbf{w}_{k+1}\} = 0$$

而 $\mathbf{u}_k$ 不是随机变量. 所以有

$$\bar{\mathbf{x}}_{k+1} = A(k)\bar{\mathbf{x}}_k + B(k)\mathbf{u}_k$$

进一步由 (1.26) 式减去 (1.27) 式得

$$\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1} = A(k)(\mathbf{x}_k - \bar{\mathbf{x}}_k) + \Gamma(k)\mathbf{w}_{k+1}$$

从而有

$$\begin{aligned} &(\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1})(\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1})^\tau \\ &= A(k)(\mathbf{x}_k - \bar{\mathbf{x}}_k)(\mathbf{x}_k - \bar{\mathbf{x}}_k)^\tau A(k)^\tau \\ &\quad + \Gamma(k)\mathbf{w}_{k+1}\mathbf{w}_{k+1}^\tau \Gamma(k)^\tau + \cdots \end{aligned}$$

注意到 $\mathbf{w}_{k+1}$ 与 $\mathbf{x}_k$ 之间的独立性,上式两端取均值即得(1.28)式.

用状态空间模型描述一个动态系统时,除状态方程外,一般地说,还要有一个观测方程(输出方程).其一般形式是

$$\mathbf{y}_k = h[\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, k] \tag{1.29}$$

其中 $\mathbf{y}_k$ 是 l 维的输出，$\mathbf{v}_k$ 是 q 维的随机噪声。

线性的情形是

$$\mathbf{y}_k = C(k)\mathbf{x}_k + D(k)\mathbf{u}_k + F(k)\mathbf{v}_k \tag{1.30}$$

其中 $C(k)$、$D(k)$ 和 $F(k)$ 分别是 $l \times n$, $l \times m$ 和 $l \times q$ 矩阵。

无噪声情形的线性状态空间模型是

$$\mathbf{x}_{k+1} = A(k)\mathbf{x}_k + B(k)\mathbf{u}_k$$
$$\mathbf{y}_k = C(k)\mathbf{x}_k + D(k)\mathbf{u}_k$$

此时，我们说该系统是确定性的(非随机的)。 如果矩阵 $A(k)$，$B(k)$，$C(k)$、$D(k)$ 均不依赖于 k,则说系统是非时变的。此时有

$$\mathbf{x}_{k+1} = A\mathbf{x}_k + B\mathbf{u}_k$$
$$\mathbf{y}_k = C\mathbf{x}_k + D\mathbf{u}_k$$

对于这个系统,如果状态初值 x_0 是已知的。则我们有

$$\mathbf{x}_1 = A\mathbf{x}_0 + B\mathbf{u}_0$$
$$\mathbf{x}_2 = A\mathbf{x}_1 + B\mathbf{u}_1 = A^2\mathbf{x}_0 + AB\mathbf{u}_0 + B\mathbf{u}_1$$
$$\mathbf{x}_3 = A^3\mathbf{x}_0 + A^2B\mathbf{u}_0 + AB\mathbf{u}_1 + B\mathbf{u}_2$$

一般地有

$$\mathbf{x}_{k+1} = A^{k+1}\mathbf{x}_0 + \sum_{i=0}^{k} A^{k-i}B\mathbf{u}_i \tag{1.31}$$

置

$$\Phi(k) = A^k \quad \Phi(0) = I$$

则 (1.31) 式可以写成

$$\mathbf{x}_{k+1} = \Phi(k+1)\mathbf{x}_0 + \sum_{i=0}^{k}\Phi(k-i)B\mathbf{u}_i \tag{1.32}$$

于是

$$\mathbf{y}_k = C\Phi(k)\mathbf{x}_0 + C\sum_{i=0}^{k-1}\Phi(k-1-i)B\mathbf{u}_i + D\mathbf{u}_k \tag{1.33}$$

由(1.32)和(1.33)两式可见,对于确定性非时变线性系统而言，在 k 时刻的状态和输出,完全被初始状态 $\mathbf{x}_0$ 和一系列的输入 $\mathbf{u}_0$, $\mathbf{u}_1$, $\cdots$, $\mathbf{u}_{k-1}$ 所确定。

第二章　非时变参数系统的辨识

本章主要介绍基本数学模型中所含的非时变未知参数的估计方法和 ARMA 模型阶数的确定方法.

§2.1　最小二乘法

1. 最小二乘格式

（1）回归系数的估计问题

考虑如下的某种预报问题所涉及的线性回归模型

$$y = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n + e$$

其中 y 是被预报量，$x_1, x_2, \cdots, x_n$ 是 n 个预报因子，e 是随机误差，$\theta_1, \theta_2, \cdots, \theta_n$ 是未知非时变参数. 设有 N 组观测资料

$$(y(i), x_1(i), x_2(i), \cdots, x_n(i)) \quad i = 1, 2, \cdots, N$$

则每组观测资料应该满足

$$y(i) = \theta_1 x_1(i) + \theta_2 x_2(i) + \cdots + \theta_n x_n(i) + \varepsilon(i) \qquad (2.1)$$

其中 $\varepsilon(i)$ 表示残差. 只要置

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix} \qquad \boldsymbol{\phi}(i) = \begin{pmatrix} x_1(i) \\ x_2(i) \\ \vdots \\ x_n(i) \end{pmatrix}$$

(2.1) 式就可以写成形式

$$y(i) = \boldsymbol{\phi}(i)^\tau \boldsymbol{\theta} + \varepsilon(i) \qquad i = 1, 2, \cdots, N$$

这是一种重要的表示形式，称之为最小二乘格式.

（2）随机差分方程参数的估计问题

考虑某动态系统的随机差分方程模型

$$y_k + a_1 y_{k-1} + \cdots + a_n y_{k-n} = b_0 u_k + b_1 u_{k-1} + \cdots$$

$$+ b_m u_{k-m} + e_k \tag{2.2}$$

其中 u_k 是输入，y_k 是输出，e_k 是随机干扰，且满足 $E\{e_k\} = 0$. 只要置

$$\theta = (a_1, a_2, \cdots, a_n, b_0, b_1, \cdots, b_m)^\tau$$

$$\phi(k) = (-y_{k-1}, -y_{k-2}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m})^\tau$$

则 (2.2) 式就可以写成

$$y_k = \phi(k)^\tau \theta + e_k$$

这又是前面所指出的最小二乘格式。

当然，还可以举出许多例子，来说明它们的数学模型可以写成上述的最小二乘格式。我们将会发现这种格式具有一定的普遍意义。

2. 最小二乘格式的参数估计

假设有

$$y_k = \phi(k)^\tau \theta + e_k \tag{2.3}$$

其中 y_k 是输出变量，θ 是未知参数向量，$\phi(k)$ 是可由观测所确定的向量，e_k 是零均值的白噪声。我们的目的是要寻求估计 θ 值的一种方法。设已有了N组观测数据：$(y_1;\ \phi(1)^\tau)$；$(y_2;\ \phi(2)^\tau)$；$\cdots$；$(y_N;\phi(N)^\tau)$这些数据中当然包括了未知参数 θ 的某些信息，我们将依据这些信息来估计 θ 的值。对于这些数据可有

$$y_i = \phi(i)^\tau \theta + \varepsilon_i \quad i = 1, 2, \cdots, N$$

其中 ε_i 表示相应于估值 θ 的残差，即

$$\varepsilon_i = \varepsilon_i(\theta) = y_i - \phi(i)^\tau \theta$$

自然希望选取 θ 的估值 $\hat{\theta}$，使所有的残差 $\varepsilon_i(\hat{\theta})(i = 1, 2, \cdots, N)$ 尽可能地小。这就引出了关于参量 θ 的如下的估计准则

$$J_N(\theta) = \sum_{k=1}^{N} \varepsilon_k(\theta)^2 \tag{2.4}$$

因为在 $J_N(\theta)$ 中包含有 $\varepsilon_k(\theta)$ 的平方项，而所求的估值 $\hat{\theta}$ 使得 $J_N(\theta)$ 为最小。所以准则 (2.4) 式称为最小平方准则或最小二乘准则。也可以用矩阵和向量的记号来表示准则 (2.4) 式，为此置

$$\mathbf{z}(N)=\begin{pmatrix}y_1\\y_2\\\vdots\\y_N\end{pmatrix}\quad H(N)=\begin{pmatrix}\boldsymbol{\phi}(1)^\tau\\\boldsymbol{\phi}(2)^\tau\\\vdots\\\boldsymbol{\phi}(N)^\tau\end{pmatrix}\quad \mathbf{E}(N)=\begin{pmatrix}\varepsilon_1(\theta)\\\varepsilon_2(\theta)\\\vdots\\\varepsilon_N(\theta)\end{pmatrix}$$

则有

$$\mathbf{z}(N)=H(N)\boldsymbol{\theta}+\mathbf{E}(N)$$

并且

$$\begin{aligned}J_N(\boldsymbol{\theta})&=\mathbf{E}(N)^\tau\mathbf{E}(N)\\&=[\mathbf{z}(N)-H(N)\boldsymbol{\theta}]^\tau[\mathbf{z}(N)-H(N)\boldsymbol{\theta}]\end{aligned}$$

为寻求 $\hat{\boldsymbol{\theta}}$ 使

$$J_N(\hat{\boldsymbol{\theta}})=\min_{\theta} J_N(\boldsymbol{\theta})$$

只需求解下述的方程组

$$\nabla_\theta J_N(\hat{\boldsymbol{\theta}})=0$$

此处 $\nabla_\theta J_N(\boldsymbol{\theta})$ 表示 $J_N(\boldsymbol{\theta})$ 的梯度.

具体地有

$$-2H(N)^\tau\mathbf{z}(N)+2H(N)^\tau H(N)\hat{\boldsymbol{\theta}}=0$$

即

$$\hat{\boldsymbol{\theta}}=[H(N)^\tau H(N)]^{-1}H(N)^\tau\mathbf{z}(N)\qquad(2.5)$$

由于 $\hat{\boldsymbol{\theta}}$ 依赖于观测数据的组数 N，所以把它记为 $\hat{\boldsymbol{\theta}}_N$ 或 $\hat{\boldsymbol{\theta}}(N)$.

如果考虑估计准则

$$J_N(\boldsymbol{\theta})=[\mathbf{z}(N)-H(N)\boldsymbol{\theta}]^\tau W(N)[\mathbf{z}(N)-H(N)\boldsymbol{\theta}]$$

其中 $W(N)$ 是某个正定的加权矩阵，$\boldsymbol{\theta}$ 的估值公式则是

$$\hat{\boldsymbol{\theta}}_N=[H(N)^\tau W(N)H(N)]^{-1}H(N)^\tau W(N)\mathbf{z}(N)\qquad(2.6)$$

3. 最小二乘估值的无偏性

我们称 (2.6) 式为参量 $\boldsymbol{\theta}$ 的加权最小二乘估值，不难证明，它具有下述的无偏性：

定理 2.1 如果 $\{e_k\}$ 是零均值的白噪声并且 $E(N)$ 与 $H(N)$ 独立，则有

$$E\{\hat{\boldsymbol{\theta}}_N\}=\boldsymbol{\theta}$$

即最小二乘估值是无偏的。

证明 只需注意到

$$\hat{\theta}_N = [H(N)^\tau W(N)H(N)]^{-1}H(N)^\tau W(N)\mathbf{z}(N)$$
$$= [H(N)^\tau W(N)H(N)]^{-1}H(N)^\tau W(N)H(N)\theta$$
$$+ [H(N)^\tau W(N)H(N)]^{-1}H(N)^\tau W(N)\mathbf{E}(N)$$
$$= \theta + [H(N)^\tau W(N)H(N)^{-1}]H(N)^\tau W(N)\mathbf{E}(N)$$

而且 $H(N)$ 与 $\mathbf{E}(N)$ 独立，所以有

$$E\{\hat{\theta}_N\} = \theta$$ ■

§2.2 最小二乘估计的递推算法

递推算法对动态系统参数的在线辨识和算法的简化具有重要的意义.

1. 简单的递推算法

考虑系统

$$y_k = \boldsymbol{\phi}(k)^\tau\theta + e_k$$

其中 y_k 是输出，e_k 是零均值的白噪声，$\boldsymbol{\phi}(k)$ 是可由观测确定的向量，θ 是未知参数向量. 前面已指出，θ 依据N组观测数据的最小二乘估计值是

$$\hat{\theta}_N = [H(N)^\tau H(N)]^{-1}H(N)^\tau\mathbf{z}(N)$$

其中

$$H(N) = \begin{pmatrix} \boldsymbol{\phi}(1)^\tau \\ \boldsymbol{\phi}(2)^\tau \\ \vdots \\ \boldsymbol{\phi}(N)^\tau \end{pmatrix} \qquad \mathbf{z}(N) = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

如果又增加了一组新的观测数据 $(y_{N+1}, \boldsymbol{\phi}(N+1)^\tau)$，通过公式

$$\hat{\theta}_{N+1} = [H(N+1)^\tau H(N+1)]^{-1}H(N+1)^\tau\mathbf{z}(N+1)$$

当然可以得出估值 $\hat{\theta}_{N+1}$. 然而这种对过去的数据完全重复一遍的算法，是繁琐的不经济的，应该寻求较简便的途径. 为此置

$$P(N) = [H(N)^{\tau}H(N)]^{-1}$$

于是

$$\begin{aligned}
P(N+1) &= [H(N+1)^{\tau}H(N+1)]^{-1} \\
&= \left[\begin{pmatrix} H(N) \\ \boldsymbol{\phi}(N+1)^{\tau} \end{pmatrix}^{\tau} \begin{pmatrix} H(N) \\ \boldsymbol{\phi}(N+1)^{\tau} \end{pmatrix}\right]^{-1} \\
&= [H(N)^{\tau}H(N) + \boldsymbol{\phi}(N+1)\boldsymbol{\phi}(N+1)^{\tau}]^{-1} \\
&= [P(N)^{-1} + \boldsymbol{\phi}(N+1)\boldsymbol{\phi}(N+1)^{\tau}]^{-1}
\end{aligned}$$

利用熟知的矩阵求逆引理，我们有

$$\begin{aligned}
P(N+1) &= P(N) - \frac{P(N)\boldsymbol{\phi}(N+1)}{1 + \boldsymbol{\phi}(N+1)^{\tau}P(N)\boldsymbol{\phi}(N+1)} \\
&\quad \cdot \boldsymbol{\phi}(N+1)^{\tau}P(N) \\
&= \{I - M(N+1)\boldsymbol{\phi}(N+1)^{\tau}\}P(N)
\end{aligned}$$

其中 I 是单位矩阵，而

$$M(N+1) = \frac{P(N)\boldsymbol{\phi}(N+1)}{1 + \boldsymbol{\phi}(N+1)^{\tau}P(N)\boldsymbol{\phi}(N+1)}$$

此时

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{N+1} &= P(N+1)H(N+1)^{\tau}\mathbf{z}(N+1) \\
&= P(N+1)\begin{pmatrix} H(N) \\ \boldsymbol{\phi}(N+1)^{\tau} \end{pmatrix}^{\tau} \begin{pmatrix} \mathbf{z}(N) \\ y_{N+1} \end{pmatrix} \\
&= [I - M(N+1)\boldsymbol{\phi}(N+1)^{\tau}]P(N)[H(N)^{\tau}\mathbf{z}(N) \\
&\quad + \boldsymbol{\phi}(N+1)y_{N+1}] \\
&= \hat{\boldsymbol{\theta}}_N - M(N+1)\boldsymbol{\phi}(N+1)^{\tau}\hat{\boldsymbol{\theta}}_N \\
&\quad + [P(N)\boldsymbol{\phi}(N+1) - M(N+1)\boldsymbol{\phi}(N+1)^{\tau} \\
&\quad \cdot P(N)\boldsymbol{\phi}(N+1)]y_{N+1}
\end{aligned}$$

但

$$\begin{aligned}
M(N+1) &+ M(N+1)\boldsymbol{\phi}(N+1)^{\tau}P(N)\boldsymbol{\phi}(N+1) \\
&= P(N)\boldsymbol{\phi}(N+1)
\end{aligned}$$

所以有

$$\hat{\boldsymbol{\theta}}_{N+1} = \hat{\boldsymbol{\theta}}_N + M(N+1)[y_{N+1} - \boldsymbol{\phi}(N+1)^{\tau}\hat{\boldsymbol{\theta}}_N]$$

总之我们证明了如下定理：

定理 2.2 系统

$$y_k = \phi(k)^\tau \theta + e_k$$

中的未知参数向量 θ 的最小二乘估值可由下述的递推公式得出

$$\hat{\theta}_{k+1} = \hat{\theta}_k + M(k+1)[y_{k+1} - \phi(k+1)^\tau \hat{\theta}_k] \qquad (2.7)$$

$$M(k+1) = \frac{P(k)\phi(k+1)}{1 + \phi(k+1)^\tau P(k)\phi(k+1)} \qquad (2.8)$$

$$P(k+1) = [I - M(k+1)\phi(k+1)^\tau]P(k) \qquad (2.9)$$

其中 I 是单位矩阵，$\hat{\theta}(k)$ 的初值 $\hat{\theta}(0)$ 可任意（但要适当地）选取，$P(k)$ 的初值 $P(0)$ 可取为 $[H(0)^\tau H(0)]^{-1}$，或

$$P(0) = \frac{1}{\varepsilon} I$$

ε 是较小的正数。

2. 遗忘因子法

前面得出的递推算法 (2.7)～(2.9) 式与最小二乘的非递推算法是等价的，所以 θ 的每一个估值都与过去的所有观测资料有关。由于过多地依赖以往数据，在实践中可能会发生数据饱和现象，从而导致新的观测数据所提供的信息，被大量的老数据所淹没，造成估值的准确性下降。为克服这种现象，并使所得的算法对于参数的某些时变特性具有一定的适应能力，可采用遗忘因子法、限定记忆法等等，这里我们介绍具有普遍意义的遗忘因子法。

仍考虑系统

$$y_k = \phi(k)^\tau \theta + e_k$$

遗忘因子法对 θ 进行估计所应用的估计准则是使

$$J_N(\theta) = \sum_{k=1}^{N} \alpha^{N-k}[y_k - \phi(k)^\tau \theta]^\tau \qquad (2.10)$$

为最小，其中 $\alpha = \beta^2$ 是遗忘因子，它满足 $0 < \alpha \leqslant 1$，α 的值愈小，对于老数据遗忘得愈快。

只要置

$$\mathbf{z}(N)=\begin{pmatrix}\beta\mathbf{z}(N-1)\\ y_N\end{pmatrix}\qquad H(N)=\begin{pmatrix}\beta H(N-1)\\ \boldsymbol{\phi}(N)^{\tau}\end{pmatrix}$$

即可把 $J_N(\boldsymbol{\theta})$ 写成

$$J_N(\boldsymbol{\theta})=[\mathbf{z}(N)-H(N)\boldsymbol{\theta}]^{\tau}[\mathbf{z}(N)-H(N)\boldsymbol{\theta}]$$

从而有

$$\hat{\boldsymbol{\theta}}_N=[H(N)^{\tau}H(N)]^{-1}H(N)^{\tau}\mathbf{z}(N) \tag{2.11}$$

令

$$P(N)=[H(N)^{\tau}H(N)]^{-1}$$

由矩阵求递定理，不难得出

$$\begin{aligned}P(N+1)&=[H(N+1)^{\tau}H(N+1)]^{-1}\\&=[\beta^2H(N)^{\tau}H(N)+\boldsymbol{\phi}(N+1)\boldsymbol{\phi}(N+1)^{\tau}]^{-1}\\&=\alpha^{-1}P(N)-\alpha^{-1}P(N)\boldsymbol{\phi}(N+1)[1+\boldsymbol{\phi}(N\\&\quad+1)^{\tau}\alpha^{-1}P(N)\boldsymbol{\phi}(N+1)]^{-1}\\&\quad\cdot\boldsymbol{\phi}(N+1)^{\tau}\alpha^{-1}P(N)\\&=\frac{1}{\alpha}\left(I-\frac{P(N)\boldsymbol{\phi}(N+1)}{\alpha+\boldsymbol{\phi}(N+1)^{\tau}P(N)\boldsymbol{\phi}(N+1)}\right.\\&\quad\left.\cdot\boldsymbol{\phi}(N+1)^{\tau}\right)P(N)\\&=\frac{1}{\alpha}[I-M(N+1)\boldsymbol{\phi}(N+1)^{\tau}]P(N)\end{aligned}$$

其中

$$M(N+1)=\frac{P(N)\boldsymbol{\phi}(N+1)}{\alpha+\boldsymbol{\phi}(N+1)^{\tau}P(N)\boldsymbol{\phi}(N+1)}$$

由此得出

$$\begin{aligned}\alpha M(N+1)&=P(N)\boldsymbol{\phi}(N+1)\\&\quad-M(N+1)\boldsymbol{\phi}(N+1)^{\tau}P(N)\boldsymbol{\phi}(N+1)\end{aligned}$$

利用上述等式，由

$$\hat{\boldsymbol{\theta}}_{N+1}=P(N+1)H(N+1)^{\tau}\mathbf{z}(N+1)$$

即可得出

$$\hat{\boldsymbol{\theta}}_{N+1}=\hat{\boldsymbol{\theta}}_N+M(N+1)[y_{N+1}-\boldsymbol{\phi}(N+1)^{\tau}\hat{\boldsymbol{\theta}}_N]$$

也就是说，我们证明了下述的定理。

定理 2.3 具有遗忘因子的最小二乘估值满足下述的递推公式

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + M(k+1)\{y_{k+1} - \boldsymbol{\phi}(k+1)^\tau \hat{\boldsymbol{\theta}}_k\} \tag{2.12}$$

$$M(k+1) = \frac{P(k)\boldsymbol{\phi}(k+1)}{\alpha + \boldsymbol{\phi}(k+1)^\tau P(k)\boldsymbol{\phi}(k+1)} \tag{2.13}$$

$$P(k+1) = \frac{1}{\alpha}[I - M(k+1)\boldsymbol{\phi}(k+1)^\tau]P(k) \tag{2.14}$$

其中 α 为遗忘因子，它满足 $0 < \alpha \leqslant 1$.

3. 关于最小二乘估值一致性的讨论

对于系统

$$y_k = \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta} + e_k$$

而言，如果噪声 e_k 不是白噪声，则未知参数的最小二乘估值可能不具有一致性，并且是有偏的.

例：考虑如下的模型

$$y_k = a y_{k-1} + b u_k + \varepsilon_k \tag{2.15}$$

的未知参数 $\boldsymbol{\theta} = (a, b)^\tau$ 的最小二乘估值，其中 $\{\varepsilon_k\}$ 和 $\{u_k\}$ 皆为零均值的广义平稳序列，并且它们之间是相互独立的. 对于系统 (2.15)，我们有

$$\hat{\boldsymbol{\theta}}_N = \begin{pmatrix} \hat{a}_N \\ \hat{b}_N \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{k=1}^N y_{k-1}^2 & \frac{1}{N}\sum_{k=1}^N u_k y_{k-1} \\ \frac{1}{N}\sum_{k=1}^N y_{k-1} u_k & \frac{1}{N}\sum_{k=1}^N u_k^2 \end{pmatrix}^{-1}$$

$$\cdot \begin{pmatrix} \frac{1}{N}\sum_{k=1}^N y_k y_{k-1} \\ \frac{1}{N}\sum_{k=1}^N y_k u_k \end{pmatrix}$$

进一步假定序列 $\{y_k\}$ 和 $\{u_k\}$ 满足遍历性质，于是当 $N \to \infty$ 时，我们有

$$\begin{pmatrix} \hat{a}_N \\ \hat{b}_N \end{pmatrix} \to \begin{pmatrix} R_{yy}(0) & R_{yu}(1) \\ R_{uy}(1) & R_{uu}(0) \end{pmatrix}^{-1} \begin{pmatrix} R_{yy}(1) \\ R_{uy}(0) \end{pmatrix} \tag{2.16}$$

其中 $R_{yy}(k)$, $R_{uu}(k)$, $R_{uy}(k)$ 分别表示序列 $\{y_k\}$, $\{u_k\}$ 的自协方差函数和互协方差函数. 如果令

$$\Delta = R_{yy}(0)R_{uu}(0) - R_{uy}(1)^2$$

则 (2.16) 式可以写成

$$\begin{pmatrix} \hat{a}_N \\ \hat{b}_N \end{pmatrix} \to \frac{1}{\Delta} \begin{pmatrix} R_{uu}(0)R_{yy}(1) - R_{yu}(1)R_{uy}(0) \\ R_{yy}(0)R_{uy}(0) - R_{yy}(1)R_{uy}(1) \end{pmatrix} \tag{2.17}$$

进一步我们假设 $\{\varepsilon_k\}$ 的协方差函数满足

$$R_{\varepsilon\varepsilon}(\tau) = 0 \qquad |\tau| \geqslant 2$$

于是由

$$y_k = a y_{k-1} + b u_k + \varepsilon_k$$

得出

$$R_{yy}(1) = aR_{yy}(0) + bR_{uy}(1) + R_{\varepsilon\varepsilon}(1)$$
$$R_{uy}(0) = aR_{uy}(1) + bR_{uu}(0)$$

把它们代入 (2.17) 式,即得

$$\begin{pmatrix} \hat{a}_N \\ \hat{b}_N \end{pmatrix} \to \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} R_{\varepsilon\varepsilon}(1)R_{uu}(0)/\Delta \\ -R_{\varepsilon\varepsilon}(1)K_{uy}(1)/\Delta \end{pmatrix}$$

这说明 $\hat{a}_N$, $\hat{b}_N$, 并不按概率或以概率 1(a. s) 收敛到 a, b, 而且也不具有无偏性.

§2.3 辅助变量法

这里,我们仍考虑系统

$$y_k = \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta} + v_k \tag{2.18}$$

的未知参量 $\boldsymbol{\theta}$ 的估值问题,但不假定 $\{v_k\}$ 为白噪声,噪声 v_k 的统计特性具有某种任意性. 为使所得到的估值 $\hat{\boldsymbol{\theta}}_N$ 具有一致性,即

$$\lim_{N\to\infty} \hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta} \qquad \text{a.s}$$

我们来介绍下述的辅助变量法.

依据观测数据, 我们把 (2.12) 式变成下述的矩阵、向量的表

示形式

$$\mathbf{z}(N) = H(N)\boldsymbol{\theta} + \mathbf{v}(N)$$

其中

$$\mathbf{z}(N) = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \qquad \mathbf{v}(N) = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} \qquad H(N) = \begin{pmatrix} \boldsymbol{\phi}(1)^\tau \\ \boldsymbol{\phi}(2)^\tau \\ \vdots \\ \boldsymbol{\phi}(N)^\tau \end{pmatrix}$$

于是由最小二乘法得到的关于参量 $\boldsymbol{\theta}$ 的估值是

$$\begin{aligned}\hat{\boldsymbol{\theta}}_N &= [H(N)^\tau H(N)]^{-1}H(N)^\tau \mathbf{z}(N) \\ &= \boldsymbol{\theta} + [H(N)^\tau H(N)]^{-1}H(N)^\tau \mathbf{v}(N) \\ &= \boldsymbol{\theta} + \left[\frac{1}{N} H(N)^\tau H(N)\right]^{-1}\left[\frac{1}{N} H(N)^\tau \mathbf{v}(N)\right]\end{aligned}$$

但

$$\frac{1}{N} H(N)^\tau H(N) = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\phi}(k)\boldsymbol{\phi}(k)^\tau$$

$$\frac{1}{N} H(N)^\tau \mathbf{v}(N) = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\phi}(k)v_k$$

所以只要相应的遍历性质满足，当 $N \to \infty$ 时就有

$$\frac{1}{N} H(N)^\tau H(N) \xrightarrow[\text{以概率 } 1]{\text{按概率}} R_{\phi\phi}(0) = E\{\boldsymbol{\phi}(N)\boldsymbol{\phi}(N)^\tau\}$$

$$\frac{1}{N} H(N)^\tau \mathbf{v}(N) \xrightarrow[\text{以概率 } 1]{\text{按概率}} R_{\phi v}(0) = E\{\boldsymbol{\phi}(N)v_N\}$$

从而

$$\hat{\boldsymbol{\theta}}_N \xrightarrow[\text{以概率 } 1]{\text{按概率}} \boldsymbol{\theta} + [R_{\phi\phi}(0)]^{-1}R_{\phi v}(0)$$

这说明，只要 $R_{\phi v}(0) \neq 0$，我们的算法就不具有一致性。

为克服可能发生的非一致性，我们必须对原有的算法进行改进，其方法之一是引入所谓辅助变量。具体地说，是把原来的估值公式改为下述的形式

$$\hat{\boldsymbol{\theta}}_N = [H^*(N)^\tau H(N)]^{-1}H^*(N)^\tau \mathbf{z}(N)$$

其中矩阵 $H^*(N)$ 必须满足：当 $N\to\infty$ 时，

1）

$$\frac{1}{N}H^*(N)^\tau H(N)\xrightarrow[\text{以概率1}]{\text{按概率}}R_{\phi^*\phi}(0)$$

而 $R_{\phi^*\phi}(0)$ 是一个非奇异矩阵；

2）

$$\frac{1}{N}H^*(N)\mathbf{v}(N)\xrightarrow[\text{以概率1}]{\text{按概率}}0$$

我们把具有上述性质的矩阵 $H^*(N)$ 称为辅助变量矩阵，而估值

$$\hat{\boldsymbol{\theta}}_N=[H^*(N)^\tau H(N)]^{-1}H^*(N)^\tau\mathbf{z}(N)$$

称为辅助变量估值．由于 $H^*(N)$ 具有性质 1）和 2），所以我们有

$$\hat{\boldsymbol{\theta}}_N=\boldsymbol{\theta}+[H^*(N)^\tau H(N)]^{-1}H^*(N)^\tau\mathbf{v}(N)$$

$$\xrightarrow[\text{以概率1}]{\text{按概率}}\boldsymbol{\theta}\qquad(\text{当 }N\to\infty)$$

即这种估值恒具有一致性．

以下我们导出递推的辅助变量算法，同时介绍一种选取辅助变量的方法．

考虑系统

$$A(q^{-1})y_k=B(q^{-1})u_k+v_k\tag{2.19}$$

其中

$$A(q^{-1})=1+a_1q^{-1}+\cdots+a_nq^{-n}$$
$$B(q^{-1})=b_0+b_1q^{-1}+\cdots+b_mq^{-m}$$

$\{v_k\}$ 是有色（相关）噪声，但 $E\{v_k\}=0$．置

$$\boldsymbol{\theta}=(a_1,a_2,\cdots,a_n,b_0,b_1,\cdots,b_m)^\tau$$

$$\boldsymbol{\phi}(k)=(-y_{k-1},\cdots,-y_{k-n},u_k,u_{k-1},\cdots,u_{k-m})^\tau$$

则可将模型 (2.19) 写为

$$y_k=\boldsymbol{\phi}(k)^\tau\boldsymbol{\theta}+v_k$$

再令

$$\mathbf{z}(N)=\begin{pmatrix}y_1\\y_2\\\vdots\\y_N\end{pmatrix}\qquad H(N)=\begin{pmatrix}\boldsymbol{\phi}(1)^\tau\\\boldsymbol{\phi}(2)^\tau\\\vdots\\\boldsymbol{\phi}(N)^\tau\end{pmatrix}$$

设已取定了辅助变量矩阵

$$H^*(N)=\begin{pmatrix}\boldsymbol{\phi}^*(1)^\tau\\\boldsymbol{\phi}^*(2)^\tau\\\vdots\\\boldsymbol{\phi}^*(N)^\tau\end{pmatrix}$$

置

$$P(k)=[H^*(k)^\tau H(k)]^{-1}$$

应用矩阵求逆引理可得

$$\begin{aligned}P(k+1)&=\left[\begin{pmatrix}H^*(k)\\\boldsymbol{\phi}^*(k+1)^\tau\end{pmatrix}^\tau\begin{pmatrix}H(k)\\\boldsymbol{\phi}(k+1)^\tau\end{pmatrix}\right]^{-1}\\&=[H^*(k)^\tau H(k)+\boldsymbol{\phi}^*(k+1)\boldsymbol{\phi}(k+1)^\tau]^{-1}\\&=P(k)-\frac{P(k)\boldsymbol{\phi}^*(k+1)}{1+\boldsymbol{\phi}(k+1)^\tau P(k)\boldsymbol{\phi}^*(k+1)}\\&\quad\cdot\boldsymbol{\phi}(k+1)^\tau P(k)\\&=[I-M(k+1)\boldsymbol{\phi}(k+1)^\tau]P(k)\end{aligned}$$

其中

$$M(k+1)=\frac{P(k)\boldsymbol{\phi}^*(k+1)}{1+\boldsymbol{\phi}(k+1)^\tau P(k)\boldsymbol{\phi}^*(k+1)}$$

由此得出

$$\begin{aligned}M(k+1)=&P(k)\boldsymbol{\phi}^*(k+1)\\&-M(k+1)\boldsymbol{\phi}(k+1)^\tau P(k)\boldsymbol{\phi}^*(k+1)\end{aligned}$$

所以有

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{k+1}&=P(k+1)H^*(k+1)^\tau\mathbf{z}(k+1)\\&=P(k+1)[H^*(k)^\tau\mathbf{z}(k)+\boldsymbol{\phi}^*(k+1)y_{k+1}]\\&=P(k)H^*(k)^\tau\mathbf{z}(k)-M(k+1)\boldsymbol{\phi}(k+1)^\tau\end{aligned}$$

$$\cdot P(k)H^*(k)\mathbf{z}(k) + [P(k)\boldsymbol{\phi}^*(k+1)$$
$$- M(k+1)\boldsymbol{\phi}(k+1)^{\tau}P(k)\boldsymbol{\phi}^*(k+1)]y_{k+1}$$
$$= \hat{\boldsymbol{\theta}}_k + M(k+1)\{y_{k+1} - \boldsymbol{\phi}(k+1)^{\tau}\hat{\boldsymbol{\theta}}_k\}$$

这就是说，我们证明了下述的定理：

定理 2.4 递推的辅助变量算法由下述各式给出

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + M(k+1)\{y_{k+1} - \boldsymbol{\phi}(k+1)^{\tau}\hat{\boldsymbol{\theta}}_k\} \tag{2.20}$$

$$M(k+1) = \frac{P(k)\boldsymbol{\phi}^*(k+1)}{1 + \boldsymbol{\phi}(k+1)^{\tau}P(k)\boldsymbol{\phi}^*(k+1)} \tag{2.21}$$

$$P(k+1) = [I - M(k+1)\boldsymbol{\phi}(k+1)^{\tau}]P(k) \tag{2.22}$$

余下的问题是怎样选取辅助变量向量 $\boldsymbol{\phi}^*(k)$. 一个自然的想法是取

$$\boldsymbol{\phi}^*(k) = [-y_{k+1}^*, -y_{k-1}^*, \cdots, -y_{k-n}^*, u_k, u_{k-1}, \cdots, u_{k-m}]$$

其中 y_k^* 满足

$$A(q^{-1})y_k^* = B(q^{-1})u_k$$

即 u_k 是原有的输入，而 y_k^* 是上述无噪声系统的输出. 这样得出的 $H^*(N)$ 既可能与 $\mathbf{v}(N)$ 独立，又可能使 $H^*(N)^{\tau}H(N)$ 非奇异，但遗憾的是 $A(q^{-1})$ 和 $B(q^{-1})$ 尚未知，所以我们不得不寻求另外的途径. 一个有效的途径是在递推算法中取

$$\boldsymbol{\phi}^*(k) = [-x_{k-1}, \cdots, -x_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}]^{\tau}$$

其中

$$x_k = \boldsymbol{\phi}^*(k)^{\tau}\hat{\boldsymbol{\theta}}_k$$

或者

$$x_k = \boldsymbol{\phi}^*(k)^{\tau}\hat{\boldsymbol{\theta}}_{k-l}$$

l 是一个小的正整数.

不难看出，这样选取的 $\boldsymbol{\phi}^*(k)$ 既可有所要求的特性，又可与 $\hat{\boldsymbol{\theta}}_k$ 相互交替地被确定.

§ 2.4 广义最小二乘法和推广的最小二乘法

1. 广义最小二乘法

广义最小二乘法通常是用来求解具有 AR 噪声的系统参数

估计问题

$$A(q^{-1})y_k = B(q^{-1})u_k + \frac{1}{D(q^{-1})}e_k \tag{2.23}$$

其中 u_k 是输入，y_k 是输出，e_k 是零均值的白噪声，而

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$
$$B(q^{-1}) = b_0 + b_1q^{-1} + \cdots + b_mq^{-m}$$
$$D(q^{-1}) = 1 + d_1q^{-1} + \cdots + d_pq^{-p}$$

也可以把 (2.23) 式写成

$$A(q^{-1})D(q^{-1})y_k = B(q^{-1})D(q^{-1})u_k + e_k$$

如果令

$$y_v(k) = D(q^{-1})y_k \quad u_v(k) = D(q^{-1})u_k$$

则有

$$A(q^{-1})y_v(k) = B(q^{-1})u_v(k) + e_k \tag{2.24}$$

系统 (2.24) 的噪声 e_k 已经是零均值的白噪声了. 如果 $\{y_v(k)\}$ 和 $\{u_v(k)\}$ 的值是可以事先确定的，则 $A(q^{-1})$ 和 $B(q^{-1})$ 的系数可以应用普通最小二乘法确定. 遗憾地是当 $D(q^{-1})$ 未知时，直接确定 $y_v(k)$ 和 $u_v(k)$ 几乎是不可能的. 然而我们却可以采取对 $A(q^{-1})$, $B(q^{-1})$ 和 $D(q^{-1})$ 逐步的依次交互确定的途径. 具体地说，我们有

1) 令 $\hat{D}(q^{-1}) = 1$.

2) 用 $\hat{D}(q^{-1})$ 经数据 y_k 和 u_k 算出

$$y_v(k) = \hat{D}(q^{-1})y_k \quad u_v(k) = \hat{D}(q^{-1})u_k$$

3) 由数据 $\{y_v(k)\}$ 和 $\{u_v(k)\}$ 得出 $A(q^{-1})$ 和 $B(q^{-1})$ 的未知参量 $\boldsymbol{\theta}$ 的最小二乘估值，用 $\hat{A}(q^{-1})$ 和 $\hat{B}(q^{-1})$ 表示 $A(q^{-1})$ 和 $B(q^{-1})$ 的相应的估计式.

4) 计算残差

$$\theta(k) = \hat{A}(q^{-1})y_k - \hat{B}(q^{-1})u_k$$

5) 增加 $D(q^{-1})$ 的阶数，考虑模型

$$D(q^{-1})\theta(k) = e_k$$

由于 e_k 是零均值的白噪声，所以上式是一个 AR 模型. 不难依据

数据 $\{\vartheta(k)\}$ 算出 $D(q^{-1})$ 中未知系数的估值，从而得出了一个新的 $\hat{D}(q^{-1})$.

反复地进行 2)～5) 各步，直到估值满足一定的终止条件为止. 在实际应用中 $D(q^{-1})$ 的阶数取得不太高就能得到较好的结果.

我们称上述的方法为张弛算法或广义最小二乘法. 由于这种算法是由两个最小二乘估值算法交互组合而成，所以不难写出它们的递推形式,为此引入记号

$$\hat{\boldsymbol{\theta}}_1 = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n, \hat{b}_0, \hat{b}_1, \cdots, \hat{b}_m)^\tau$$

$$\hat{\boldsymbol{\theta}}_2 = (\hat{d}_1, \hat{d}_2, \cdots, \hat{d}_p)^\tau$$

$$y_v(k) = y_k + \hat{d}_1 y_{k-1} + \cdots + \hat{d}_p y_{k-p}$$

$$u_v(k) = u_k + \hat{d}_1 u_{k-1} + \cdots + \hat{d}_p u_{k-p}$$

$$\boldsymbol{\phi}_1(k) = [-y_v(k-1), \cdots, -y_v(k-n), u_v(k), \cdots, u_v(k-m)]^\tau$$

$$\boldsymbol{\phi}_2(k) = [-\vartheta(k-1), \cdots, -\vartheta(k-p)]^\tau$$

于是两组交互使用的递推算法可以写成下述形式

$$\hat{\boldsymbol{\theta}}_1(k+1) = \hat{\boldsymbol{\theta}}_1(k) + M_1(k+1)\{y_v(k+1) - \boldsymbol{\phi}_1(k+1)^\tau \hat{\boldsymbol{\theta}}_1(k)\} \quad (2.25)$$

$$M_1(k+1) = \frac{P_1(k)\boldsymbol{\phi}_1(k+1)}{1 + \boldsymbol{\phi}_1(k+1)^\tau P_1(k)\boldsymbol{\phi}_1(k+1)} \quad (2.26)$$

$$P_1(k+1) = [I - M_1(k+1)\boldsymbol{\phi}_1(k+1)^\tau]P_1(k) \quad (2.27)$$

$$\hat{\boldsymbol{\theta}}_2(k+1) = \hat{\boldsymbol{\theta}}_2(k) + M_2(k+1)\{\vartheta(k+1) - \boldsymbol{\phi}_2(k+1)^\tau \hat{\boldsymbol{\theta}}_2(k)\} \quad (2.28)$$

$$M_2(k+1) = \frac{P_2(k)\boldsymbol{\phi}_2(k+1)}{1 + \boldsymbol{\phi}_2(k+1)^\tau P_2(k)\boldsymbol{\phi}_2(k+1)} \quad (2.29)$$

$$P_2(k+1) = [I - M_2(k+1)\boldsymbol{\phi}_2(k+1)^\tau]P_2(k) \quad (2.30)$$

此外还有

$$y_v(k) = y_k + \hat{d}_1 y_{k-1} + \cdots + \hat{d}_p y_{k-p} \quad (2.31)$$

$$u_v(k) = u_k + \hat{d}_1 u_{k-1} + \cdots + \hat{d}_p u_{k-p} \quad (2.32)$$

$$\theta(k) = \hat{A}(q^{-1})y_k - \hat{B}(q^{-1})u_k \tag{2.33}$$

2. 推广的最小二乘法(增广矩阵法)

推广的最小二乘法将用来求解如下具有 MA 噪声的系统参数估计问题:

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k \tag{2.34}$$

其中 $\{e_k\}$ 是零均值的白噪声,而

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$
$$B(q^{-1}) = b_0 + b_1q^{-1} + \cdots + b_mq^{-m}$$
$$C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_pq^{-p}$$

如果置

$$\boldsymbol{\phi}(k) = [-y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}, e_{k-1}, \cdots, e_{k-p}]^\tau$$
$$\boldsymbol{\theta} = [a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, c_2, \cdots, c_p]^\tau$$

则模型 (2.34) 可以写成如下形式

$$y_k = \boldsymbol{\phi}(k)^\tau\boldsymbol{\theta} + e_k$$

如果 $e_{k-1}, \cdots, e_{k-p}$ 是可以确定的，则对于 $\boldsymbol{\theta}$ 的递推估值算法可以写成

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + M(k+1)\{y_{k+1} - \boldsymbol{\phi}(k+1)^\tau\hat{\boldsymbol{\theta}}_k\}$$
$$M(k+1) = \frac{P(k)\boldsymbol{\phi}(k+1)}{1 + \boldsymbol{\phi}(k+1)^\tau P(k)\boldsymbol{\phi}(k+1)}$$
$$P(k+1) = [I - M(k+1)\boldsymbol{\phi}(k+1)^\tau]P(k)$$

然而在实际中,$\{e_k\}$ 是不能经测量而确定的，所以上述算法是不能直接应用的,从而应该寻求关于 e_k 的一种估值方法。当 $\hat{\theta}_{k-1}$ 已接近于 θ 的真值时,不难看出

$$\hat{\varepsilon}_a(k) = y_k - \boldsymbol{\phi}_a(k)^\tau\hat{\boldsymbol{\theta}}_{(k-1)}$$

可以作为 e_k 的一种估值,其中

$$\boldsymbol{\phi}_a(k) = [-y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}, \hat{\varepsilon}_a(k-1), \cdots, \hat{\varepsilon}_a(k-p)]^\tau$$

于是所求的递推算法是

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + M(k+1)\hat{\varepsilon}_a(k+1) \tag{2.35}$$

$$M(k+1) = \frac{P(k)\boldsymbol{\phi}_a(k+1)}{1+\boldsymbol{\phi}_a(k+1)^{\tau}P(k)\boldsymbol{\phi}_a(k+1)} \tag{2.36}$$

$$P(k+1) = [I - M(k+1)\boldsymbol{\phi}_a(k+1)^{\tau}]P(k) \tag{2.37}$$

$$\hat{\varepsilon}_a(k+1) = y_{k+1} - \boldsymbol{\phi}_a(k+1)^{\tau}\hat{\boldsymbol{\theta}}_k \tag{2.38}$$

上述算法称为增广矩阵法或推广的递推最小二乘法，这种算法在实际中已得到了广泛的应用。

§2.5 预报误差估值与极大似然法

1. 预报误差估值

这里，我们来考虑用预报误差方程所描述系统的参数估计问题

$$\mathbf{y}_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\theta}, k] + \varepsilon_k \tag{2.39}$$

其中 $\mathbf{y}_k$ 是 n 维输出，$\mathbf{u}_k$ 是 p 维输入，$\boldsymbol{\theta}$ 是 m 维未知噪声，ε_k 是新息序列，特别情形可以是零均值的白噪声。

$$\mathbf{Y}_{k-1} = \{\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1}\}$$

$$\mathbf{U}_k = \{\mathbf{u}_0, \mathbf{u}_1, \cdots, \mathbf{u}_k\}$$

对于已确定的 $\boldsymbol{\theta}$ 值 $(\boldsymbol{\theta} = \hat{\boldsymbol{\theta}})$，相应的预报误差为

$$\mathbf{w}_k(\hat{\boldsymbol{\theta}}) = \mathbf{y}_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}, k]$$

$\{\mathbf{w}_k(\hat{\boldsymbol{\theta}})\}$ 的样本协方差为

$$D(\hat{\boldsymbol{\theta}}) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{w}_k(\hat{\boldsymbol{\theta}})\mathbf{w}_k(\hat{\boldsymbol{\theta}})^{\tau}$$

显然一个模型的好的参数估值应该具有较小的预报误差，这相当于要求 $D(\hat{\boldsymbol{\theta}})$ 的某种正标量函数的值比较小。通常可采用的正标量函数有

$$J_1(\hat{\boldsymbol{\theta}}) = \mathrm{tr}\{WD(\hat{\boldsymbol{\theta}})\}$$

其中W是一个正定矩阵，tr 表示矩阵的迹数。以及

$$J_2(\hat{\boldsymbol{\theta}}) = \log \det D(\hat{\boldsymbol{\theta}})$$

对 J_1 或 J_2 这样的标量函数的极小化所得出的关于参量 $\boldsymbol{\theta}$ 的

估值，称为预报误差估值.

2. 极大似然法

如果我们假定模型(2.39)中的噪声序列$\{\boldsymbol{\varepsilon}_k\}$是独立的随机变量序列，并且输入$\{\mathbf{u}_k\}$与$\{\varepsilon_k\}$是相互独立的. 则应用贝叶斯(Bayes)公式可以得出似然函数的表达式如下:

$$p\{\mathbf{Y}_N|\mathbf{U}_N,\boldsymbol{\theta}\}=\prod_{k=1}^{N}p\{\mathbf{y}_k|\mathbf{Y}_{k+1},\mathbf{U}_k,\boldsymbol{\theta}\}$$

此处$p\{\cdot|\cdots\}$表示条件分布密度. 由模型(2.39)不难看出，在$\mathbf{Y}_{k-1}$，$\mathbf{U}_k$已知的条件下，$\mathbf{y}_k$的条件分布被ε_k的分布所决定，并且有

$$p\{\mathbf{y}_k|\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}\}=p_{\varepsilon_k}\{\mathbf{w}_k(\boldsymbol{\theta})|\boldsymbol{\theta}\}\left|\det\left(\frac{\partial\boldsymbol{\varepsilon}_k}{\partial\mathbf{y}_k}\right)\right|$$

其中

$$\left|\det\left(\frac{\partial\varepsilon_k}{\partial\mathbf{y}_k}\right)\right|=1$$

$$\mathbf{w}_k(\theta)=\mathbf{y}_k-f[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta},k]$$

$p_{\varepsilon_k}(\cdot|\theta)$表示$\varepsilon_k$的条件密度函数，于是我们得到了似然函数的表达式如下

$$p\{\mathbf{Y}_N|\mathbf{U}_N,\boldsymbol{\theta}\}=\prod_{k=1}^{N}P_{\varepsilon_k}\{\mathbf{w}_k(\boldsymbol{\theta})|\boldsymbol{\theta}\}$$

如果参量$\boldsymbol{\theta}$的估值$\hat{\boldsymbol{\theta}}$满足

$$P\{\mathbf{Y}_N|\mathbf{U}_N,\hat{\boldsymbol{\theta}}\}=\max_{\theta}P\{\mathbf{Y}_N|\mathbf{U}_N,\boldsymbol{\theta}\}$$

就说$\hat{\boldsymbol{\theta}}$是$\boldsymbol{\theta}$的极大似然估计.

进一步，如果$\{\boldsymbol{\varepsilon}_k\}$是独立正态序列，并且$E\{\boldsymbol{\varepsilon}_k\}=0$，$\mathrm{var}\{\varepsilon_k\}=\Sigma$. 则前述的似然函数可具体的写为

$$P(\mathbf{Y}_N|\mathbf{U}_N,\ \boldsymbol{\theta})$$

$$=\prod_{k=1}^{N}[(2\pi)^n\det\Sigma]^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{w}_k^{\tau}\Sigma^{-1}\mathbf{w}_k\right\}$$

$$= [(2\pi)^n \det\Sigma]^{-\frac{N}{2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{N} \mathbf{w}_k^\tau \Sigma^{-1}\mathbf{w}_k\right\} \quad (2.40)$$

其中 n 是 ε_k 的维数。

以下分两种情形进行分析：

1）Σ已知

考虑前述的预报误差估计的准则函数 $J_1(\theta)$. 取 $W = \Sigma^{-1}$ 有

$$J_1(\theta) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{w}_k(\theta)^\tau \Sigma^{-1}\mathbf{w}_k(\theta) = \mathrm{tr}\{\Sigma^{-1}D(\theta)\}$$

其中

$$D(\theta) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{w}_k(\theta) w_k(\theta)^\tau$$

由此可见，在Σ已知的情况下，极大化似然函数（2.40），等价于极小化准则函数 $J_1(\theta)$. 这说明，在Σ已知的条件下，θ 的极大似然估计就是在加权矩阵 $W = \Sigma^{-1}$ 的条件下以 $J_1(\theta)$ 为估计准则的预报误差估计。

2）Σ未知

由（2.40）式可得

$$J(\theta, \Sigma) = \frac{1}{2} nN \log 2\pi + \frac{N}{2} \log\det\Sigma + \frac{1}{2}\sum_{k=1}^{N} \mathbf{w}_k^\tau \Sigma^{-1}\mathbf{w}_k$$

此处 $J(\theta, \Sigma) = -\log P(\mathbf{Y}_N|\mathbf{U}_N, \theta)$. 由此可见，极大化似然函数 $P(\mathbf{Y}_N|\mathbf{U}_N, \theta)$ 等价于极小化函数 $J(\theta, \Sigma)$.

对 $J(\theta, \Sigma)$ 求关于Σ的导数，注意到

$$\frac{\partial}{\partial\Sigma}\log\det\Sigma = \Sigma^{-1}$$

即得

$$\frac{\partial J}{\partial\Sigma} = \frac{N}{2}\Sigma^{-1} - \frac{1}{2}\Sigma^{-1}\left(\sum_{k=1}^{N} \mathbf{w}_k \mathbf{w}_k^\tau\right)\Sigma^{-1}$$

令其为 0，即可得出

$$\Sigma = \frac{1}{N}\sum_{k=1}^{N} \mathbf{w}_k(\theta) W_k(\theta)^\tau = D(\theta)$$

把这样的 $\Sigma = D(\theta)$ 代入 $J(\theta, \Sigma)$ 中得

$$\begin{aligned} J_c(\theta) &= J(\theta, D(\theta)) \\ &= \frac{1}{2} nN \log 2\pi + \frac{N}{2} \log \det D(\theta) \\ &\quad + \frac{1}{2}\sum_{k=1}^{N} w_k^\tau D(\theta)^{-1} \mathbf{w}_k \\ &= \frac{1}{2} nN(\log 2\pi + 1) + \frac{N}{2} \log \det D(\theta) \end{aligned}$$

由此可见,关于 θ 极大化 $J_c(\theta)$ 等价于极小化准则函数

$$J_2(\theta) = \log \det D(\theta)$$

这说明在 Σ 未知时，极大似然法等价于以 $J_2(\theta)$ 为准则的预报误差估计方法.

3. 递推极大似然估计

我们考虑关于模型

$$A(q^{-1}) y_k = B(q^{-1}) u_k + C(q^{-1}) e_k \tag{2.41}$$

的未知参数的估计问题,其中

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + \cdots + b_m q^{-m} \\ C(q^{-1}) &= 1 + c_1 q^{-1} + \cdots + c_p q^{-p} \end{aligned}$$

置

$$\begin{aligned} \theta &= (a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, \cdots, c_p)^\tau \\ \phi(k) &= (-y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}, \\ &\qquad e_{k-1}, \cdots, e_{k-p})^\tau \end{aligned}$$

则 (2.41) 式可以写成

$$y_k = \phi(k)^\tau \theta + e_k$$

用 $\hat{y}(k|\theta)$ 表示 y_k 的向前一步的预报估值，$\varepsilon(k, \theta)$ 表示预报误差,即

$$\varepsilon(k, \theta) = y_k - \hat{y}(k|\theta)$$

并用 $\varepsilon(k-1, \theta)$, $\varepsilon(k-2, \theta)$, $\cdots$, $\varepsilon(k-P, \theta)$ 做为 e_{k-1}, e_{k-2}, $\cdots$, e_{k-p} 的估值,于是有

$$\begin{aligned}
\hat{y}(k|\theta) = & -a_1 y_{k-1} - a_2 y_{k-2} - \cdots - a_n y_{k-n} + b_0 u_k \\
& + b_1 u_{k-1} + \cdots + b_m u_{k-m} + c_1 \varepsilon(k-1, \theta) \\
& + \cdots + c_p \varepsilon(k-p, \theta) \\
= & \; c_1 y_{k-1} + \cdots + c_p y_{k-p} - c_1 \hat{y}(k-1|\theta) \\
& - \cdots - c_p \hat{y}(k-p|\theta) - a_1 y_{k-1} \\
& - \cdots - a_n y_{k-n} + b_0 u_k + \cdots + b_m u_{k-m}
\end{aligned}$$

当然直接由 (2.41) 式出发,也可以把 $\varepsilon(k, \theta)$ 写成下述形式

$$\varepsilon(k, \theta) = C(q^{-1})^{-1}[A(q^{-1})y_k - B(q^{-1})u_k]$$

注意到 y_k 和 e_k 皆是一维的. 按前面所指出的,在噪声 e_k 正态的假设下,极大似然估计等价于选取 θ 的估值 $\hat{\theta}$, 它满足

$$J_N(\hat{\theta}) = \min_{\theta} J_N(\theta)$$

其中

$$J_N(\theta) = \sum_{k=1}^{N} \varepsilon(k, \theta)^2$$

但 $\varepsilon(k, \theta)$ 一般的说,不是 θ 的线性函数,所以我们仅能导出关于 θ 的估值的近似极大似然递推算法. 为此考虑 $\varepsilon(k, \theta)$ 的泰勒 (Taylor) 展开式

$$\varepsilon(k, \theta) = \varepsilon(k, \hat{\theta}) + \left(\frac{\partial \varepsilon(k, \hat{\theta})}{\partial \theta}\right)^{\tau} (\theta - \hat{\theta}) + \alpha$$

其中 α 是余项, $\dfrac{\partial \varepsilon(k, \hat{\theta})}{\partial \theta}$ 是 $\varepsilon(k,\theta)$ 关于 θ 的梯度在 $\hat{\theta}$ 处的值. 为简化符号令

$$\phi(k, \theta) = -\frac{\partial \varepsilon(k, \theta)}{\partial \theta}$$

并置

$$\begin{aligned}
\phi(k, \theta) = \{ & -y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m} \\
& \varepsilon(k-1, \theta), \cdots, \varepsilon(k-p, \theta)\}^{\tau}
\end{aligned}$$

注意到

$$\frac{\partial\varepsilon(k,\theta)}{\partial a_i}=C(q^{-1})\{q^{-i}y_k\}=C(q^{-1})^{-1}y_{k-i}$$

$$\frac{\partial\varepsilon(k,\theta)}{\partial b_i}=-C(q^{-1})^{-1}\{q^{-i}u_k\}=-C(q^{-1})^{-1}u_{k-i}$$

$$\frac{\partial\varepsilon(k,\theta)}{\partial c_i}=-q^{-i}C(q^{-1})^{-2}\{A(q^{-1})y_k-B(q^{-1})u_k\}$$

$$=-C(q^{-1})\varepsilon(k-i,\theta)$$

所以有

$$\psi(k,\theta)=C(q^{-1})^{-1}\phi(k,\theta)$$

即

$$C(q^{-1})\psi(k,\theta)=\phi(k,\theta)$$

或

$$\psi(k,\theta)+c_1\psi(k-1,\theta)+\cdots+c_p\psi(k-p,\theta)=\phi(k,\theta)$$

用 $\theta(N)$ 表示 θ 的第 N 次估值，于是有

$$\nabla_\theta J_N(\hat{\theta}(N))=\frac{\partial}{\partial\theta}J_N(\theta)|_{\theta=\theta(N)}=0.$$

我们可以对 $J_N(\theta)$ 取如下的近似

$$J_N(\theta)\approx(\theta-\hat{\theta}(N))^\tau P(N)^{-1}(\theta-\hat{\theta}(N))+\beta_N$$

其中

$P(N)^{-1}=\frac{\partial^2}{\partial\theta^2}J_N(\theta)|_{\theta=\theta(N)}$，即它是 $J_N(\theta)$ 的二阶偏导数矩阵在 $\hat{\theta}(N)$ 处的值．令 $\Delta=\theta-\hat{\theta}(N)$，进一步我们有

$$\begin{aligned}J_{N+1}(\theta)&=\sum_{k=1}^{N+1}\varepsilon(k,\theta)^2=J_N(\theta)+\varepsilon(k+1,\theta)^2\\&\approx\Delta^\tau P(N)^{-1}\Delta+\beta_N+[\varepsilon(k+1,\hat{\theta}(N)\\&\quad-\psi(N+1)\hat{\theta}(N))^\tau\Delta]^2\\&=\Delta^\tau[P(N)^{-1}+\psi(N+1,\hat{\theta}(N))\\&\quad\cdot\psi(N+1,\hat{\theta}(N))^\tau]\Delta\\&\quad-2\varepsilon(N+1,\hat{\theta}(N))\psi(N+1,\hat{\theta}(N))^\tau\Delta+\beta_N'\end{aligned}\tag{2.42}$$

其中

$$\beta'_N = \beta_N + \varepsilon(N+1, \hat{\theta}(N))^2$$

在(2.42)式中配完全平方得

$$J_{N+1}(\theta) \approx (\Delta - \gamma_{N+1})^\tau P(N+1)^{-1}(\Delta - \gamma_{N+1}) + \beta_{N+1} \quad (2.43)$$

其中

$$P(N+1) = [P(N)^{-1} + \phi(N+1, \hat{\theta}(N))\phi(N+1, \theta(N))^\tau]^{-1} \quad (2.44)$$

$$\gamma_{N+1} = P(N+1)\phi(N+1, \theta(N))\varepsilon(N+1, \hat{\theta}(N)) \quad (2.45)$$

$$\beta_{N+1} = \beta'_N - \gamma^\tau_{N+1}P(N+1)^{-1}\gamma_{N+1}$$

对等式(2.44)应用矩阵求逆引理得

$$P(N+1) = [I - M(N+1)\phi(N+1), \hat{\theta}(N)^\tau]P(N)$$

其中

$$M(N+1) = \frac{P(N)\phi(N+1, \hat{\theta}(N))}{1 + \phi(N+1, \hat{\theta}(N))^\tau P(N)\phi(N+1, \hat{\theta}(N))}$$

从而有

$$\begin{aligned}
\gamma_{N+1} &= [P(N)\phi(N+1, \hat{\theta}(N)) - M(N+1) \\
&\quad \cdot \phi(N+1, \hat{\theta}(N))^\tau P(N)\phi(N \\
&\quad + 1, \hat{\theta}(N))]\varepsilon(N+1, \hat{\theta}(N)) \\
&= [P(N)\phi(N+1, \hat{\theta}(N)) - M(N+1) \\
&\quad \cdot (1 + \phi(N+1, \hat{\theta}(N))^\tau P(N)\phi(N+1, \hat{\theta}(N))) \\
&\quad + M(N+1)]\varepsilon(N+1, \hat{\theta}(N)) \\
&= M(N+1)\varepsilon(N+1, \hat{\theta}(N))
\end{aligned}$$

由(2.43)式看出，当取 $\Delta = \gamma_{N+1}$ 时可使 $J_{N+1}(\theta)$ 达到极小。于是若 $\hat{\theta}(N)$ 已使 $J_N(\theta)$ 达到极小，则

$$\begin{aligned}
\hat{\theta}(N+1) &= \hat{\theta}(N) + \gamma_{N+1} \\
&= \hat{\theta}(N) + M(N+1)\varepsilon(N+1, \hat{\theta}(N))
\end{aligned}$$

可使 $J_{N+1}(\theta)$ 达到极小。对于 $\varepsilon(N+1, \hat{\theta}(N)$ 可取如下的近似

$$\hat{\varepsilon}(N+1) = y_{N+1} - \hat{\phi}(N+1)^\tau\hat{\theta}(N)$$

其中

$$\hat{\phi}(N) = [-y_{N-1}, \cdots, -y_{N-n}, u_N, u_{N-1}, \cdots,$$

$$u_{N-m}, \hat{\varepsilon}(N-1), \cdots, \hat{\varepsilon}(N-P)]^{\tau}$$

总之,我们得出了如下的一组递推算法

$$\hat{\theta}(k+1) = \hat{\theta}(k) + M(k+1)\{y_{k+1} - \hat{\phi}(k+1)^{\tau}\theta(k)\} \tag{2.46}$$

$$M(k+1) = \frac{P(k)\phi(k+1)}{1+\phi(k+1)^{\tau}P(k)\phi(k+1)} \tag{2.47}$$

$$P(k+1) = [I - M(k+1)\phi(k+1)^{\tau}]P(k) \tag{2.48}$$

$$\hat{\phi}(k) = [-y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}, \hat{\varepsilon}(k-1), \cdots, \hat{\varepsilon}(k-p)]^{\tau} \tag{2.49}$$

$$\hat{\varepsilon}(k) = y_k - \hat{\phi}(k)^{\tau}\hat{\theta}(k-1) \tag{2.50}$$

$$\phi(k) + \hat{c}_1\phi(k-1) + \cdots + \hat{c}_p\phi(k-p) = \hat{\phi}(k) \tag{2.51}$$

其中 $\hat{c}_1, \cdots, \hat{c}_p$ 表示 $c_1, \cdots, c_p$ 的第 $k-1$ 次估值.

§2.6 随机逼近算法

考虑一个简单的例子,设观测模型是

$$y_k = \theta + v_k \tag{2.52}$$

其中 θ 是待估参数, v_k 是随机噪声,它满足

$$E\{v_k\} = 0, \quad E\{v_k, v_l\} = \sigma^2\delta_{k,l}$$

$$\delta_{k,l} = \begin{cases} 0 & k \neq l \\ 1 & k = l \end{cases}$$

如果已有了 n 个观测数据

$$y_1, y_2, \cdots, y_n$$

于是依据这些数据对 θ 的估值 $\hat{\theta}_n$ 可以写成

$$\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} y_k \tag{2.53}$$

假定又有了一个新的观测数据 y_{n+1}, 则 θ 的估值 $\hat{\theta}_{n+1}$ 又可写成

$$\hat{\theta}_{n+1} = \frac{1}{n+1}\sum_{k=1}^{n+1} y_k \tag{2.54}$$

不难看出, $\hat{\theta}_{n+1}$, $\hat{\theta}_n$ 和 y_{n+1} 之间具有如下的关系

$$\hat{\theta}_{n+1} = \frac{1}{n+1}\sum_{k=1}^{n+1} y_k = \frac{n}{n+1}\cdot\frac{1}{n}\sum_{k=1}^{n} y_k + \frac{1}{n+1} y_{n+1}$$

$$= \frac{1}{n+1}\hat{\theta}_n + \frac{1}{n+1} y_{n+1}$$

$$= \hat{\theta}_n - \frac{1}{n+1}\hat{\theta}_n + \frac{1}{n+1} y_{n+1}$$

$$= \hat{\theta}_n + \frac{1}{n+1}(y_{n+1} - \hat{\theta}_n)$$

其实我们得到了由 $\hat{\theta}_{k-1}$ 和 y_k 计算 $\hat{\theta}_k$ 的递推公式

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{1}{k}(y_k - \hat{\theta}_{k-1})$$

如果令 $\gamma_k = \frac{1}{k}$，显然 γ_k 满足

$$\lim_{k\to\infty}\gamma_k = 0 \quad \sum_{k=1}^{\infty}\gamma_k = +\infty \quad \sum_{k=1}^{\infty}\gamma_k^2 < +\infty$$

并且

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \gamma_k(y_k - \hat{\theta}_{k-1}) \tag{2.55}$$

由大数定律可知，在适当的条件下估值序列 $\{\hat{\theta}_k\}$ 必按概率或以概率 1 收敛到 θ。

罗宾斯 (Robbins) 和门罗 (Monro) 把上述形式的递推算法应用于求解下述类型的问题：设函数 $g(x)$ 在区间 $[a, b]$上有界，并且有唯一的零点 θ, $\theta\in[a,b]$，如何确定 θ 的估值?

设 $g(x)$ 是可以量测的，其观测方程为

$$y_k = g(x) + v_k$$

其中 v_k 是第 k 次的观测噪声，假定它满足

$$E\{v_k\} = 0 \quad E\{v_k, v_l\} = \sigma^2\delta_{kl}$$

罗宾斯和门罗引出了关于 θ 的估值的递推算法如下

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_k y_k \tag{2.56}$$

并且证明了，如果数列 $\{\gamma_k\}$ 满足

1) $\lim_{k\to\infty}\gamma_k = 0$;

2）$\sum_{k=1}^{\infty} \gamma_k = +\infty$；

3）$\sum_{k=1}^{\infty} \gamma_k^2 < +\infty$.

则必有

$$\lim_{k \to \infty} E\{(\hat{\theta}_k - \theta)^2\} = 0$$

以及

$$\lim_{k \to \infty} \hat{\theta}_k = \theta \quad \text{a.s}$$

我们把上述的递推算法称为随机逼近算法[11]. 基弗 (Kiefer) 和沃尔夫威茨 (Wolfowitz) 在他们的工作[12]中提出了另一类随机逼近算法,他们考虑了 $[a, b]$ 上的函数 $W(x)$, 企图寻求 $W(x)$的极值点 θ. 设 $W(x)$ 之值是可以量测的,其观测方程是

$$Z(x) = W(x) + v$$

其中噪声 v 满足

$$E\{v\} = 0 \quad E\{v^2\} = \sigma^2 < +\infty$$

他们证明了,如果数列 $\{\gamma_k\}$ 和 $\{c_k\}$ 满足

1）$\lim_{k \to \infty} \gamma_k = 0$, $\lim_{k \to \infty} c_k = 0$;

2）$\sum_{k=1}^{\infty} \gamma_k = +\infty$;

3）$\sum_{k=1}^{\infty} \left(\frac{\gamma_k}{c_k}\right)^2 < +\infty$.

则由递推算法

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_k \frac{Z(\hat{\theta}_k + c_k) - Z(\hat{\theta}_k - c_k)}{2c_k} \qquad (2.57)$$

所得出的序列 $\{\hat{\theta}_k\}$ 必以概率 1 收敛到 θ.

$$\theta \in \left\{\theta : \frac{dW(\theta)}{d\theta} = 0\right\}$$

同时还有

$$\lim_{k \to +\infty} E\{(\hat{\theta}_k - \theta)^2\} = 0$$

德沃雷茨基（Dvoretzky）提出，任何一种随机逼近方法都可以看成是一种通常的无误差逐次逼近法并迭加一个随机噪声分量而构成的。这种观点的好处是分离出随机分量，使得它能够与确定型部分隔离开来处理。以递推算法 2.57 为例：

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_k y_k$$

由于

$$y_k = g(\hat{\theta}_k) + v_k$$

所以

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_k g(\hat{\theta}_k) + \gamma_k v_k \tag{2.58}$$

只要置

$$T(\hat{\theta}_k) = \hat{\theta}_k + \gamma_k g(\hat{\theta}_k) \quad y_k = \gamma_k v_k$$

则 (2.58) 式可以写成

$$\hat{\theta}_{k+1} = T(\hat{\theta}_k) + y_k \tag{2.59}$$

其中，$T(\hat{\theta}_k)$ 表示算法的确定性部分，一般的它是关于自变量的一个可测变换，y_k 表示随机部分。

如果用 θ 表示参量的真值，自然希望我们的算法 (2.59) 中的确定性部分 $T(\hat{\theta}_k)$ 具有性质

$$|T(\hat{\theta}_k) - \theta| \leqslant \alpha_k$$

而 $\lim\limits_{k\to\infty}\alpha_k = 0$，以及

$$|T(\hat{\theta}_k) - \theta| \leqslant |\hat{\theta}_k - \theta| - \gamma_k$$

这个不等式意味着递推算法每前进一步，$\hat{\theta}_k - \theta$ 都能有所减小。虽然当 $\hat{\theta}_k$ 趋近于 θ 时，γ_k 应该是很小的，但是并不希望 γ_k 太快地趋近于零，这是因为要避免在达到目标最优点之前递推过程过早停止。所以我们要求

$$\sum_{k=1}^{\infty} \gamma_k = \infty$$

实际上在开始阶段，不一定总要使步长缩减。德沃雷茨基给出了下面的较弱条件

$$|T(\hat{\theta}_k) - \theta| \leqslant (1 + \beta_k)|\hat{\theta}_k - \theta| - \gamma_k \tag{2.60}$$

其中

$$\sum_{k=1}^{\infty} \beta_k < +\infty$$

德沃雷茨基证明了下述关于较一般随机逼近算法的收敛性定理。

定理 2.5（德沃雷茨基定理） 设 α_k, β_k 和 $\gamma_k(k=1,2,\cdots)$ 皆为非负实数，且满足

$$\lim_{k\to\infty} \alpha_k = 0$$

$$\sum_{k=1}^{\infty} \beta_k < +\infty$$

$$\sum_{k=1}^{\infty} \gamma_k = +\infty$$

设 θ 是实数，$T_k(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ 是一个可测的变换，并且对一切 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ 满足

$$|T_k(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) - \theta| \leqslant \max\{\alpha_k, (1+\beta_k)|\hat{\theta}_k - \theta| - \gamma_k\}$$

又设 y_k 是一个随机变量，它满足

$$\sum_{k=1}^{\infty} E\{y_k^2\} < +\infty \quad E\{y_k\} = 0$$

以及

$$E\{\hat{\theta}_1^2\} < +\infty$$

则由递推算法

$$\hat{\theta}_{k+1} = T_k(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) + y_k \tag{2.61}$$

所确定的序列 $\{\hat{\theta}_k\}$ 必满足

$$P\{\lim_{k\to\infty} \hat{\theta}_k = \theta\} = 1$$

以及

$$\lim_{k\to+\infty} E\{(\hat{\theta}_k - \theta)^2\} = 0$$

本定理的证明请参看文献[13]。我们可以把上述定理推广到一般情形，于是有如下定理：

定理 2.6 设 $\boldsymbol{\theta}_k$ 是赋范线性空间 Θ 中的向量，$T_k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots,$

$\hat{\boldsymbol{\theta}}_k$) 是从 $\Theta^k = \Theta \times \Theta \times \cdots \times \Theta$ 到 Θ 的可测变换，它满足

$$\|T_k(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \cdots, \hat{\boldsymbol{\theta}}_k) - \boldsymbol{\theta}\| \leqslant (1 - \gamma_k)\|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\|$$

此处 $\boldsymbol{\theta} \in \Theta$，并且

$$1 - \gamma_k > 0 \quad \prod_{k=1}^{\infty} (1 - \gamma_k) = 0 \quad \sum_{k=1}^{\infty} \gamma_k^2 < +\infty$$

又设

$$\hat{\boldsymbol{\theta}}_{k+1} = T_k(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \cdots, \hat{\boldsymbol{\theta}}_k) + \mathbf{y}_k$$

$\mathbf{y}_k$ 是在 Θ 中取值的随机元. 如果对于 Θ 上的任何可测变换 $\phi(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2, \cdots, \hat{\boldsymbol{\theta}}_n)$，皆有

$$E\{\|\phi(\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_k) + \mathbf{y}_k\|^2\} < E\{\|\phi(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)\|^2\} + E\{\|\mathbf{y}_k\|^2\}$$

以及

$$E\{\|\hat{\boldsymbol{\theta}}_1\|^2\} + \sum_{k=1}^{\infty} E\{\|\mathbf{y}_k\|^2\} < +\infty$$

则有

$$P\{\lim_{k\to\infty} \|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\| = 0\} = 1$$

$$\lim_{k\to\infty} E\{\|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\|^2\} = 0$$

我们可以把上述关于随机逼近的思想，应用于求解系统

$$y_k = \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta} + e_k$$

的参数估计问题. 这里，我们所寻求的是估计准则

$$J[\boldsymbol{\theta}] = E\{(y_k - \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta})^2\}$$

的最小值点. 即函数 $E\{\boldsymbol{\phi}(k)(y_k - \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta})\}$ 的零点. 由于

$$\boldsymbol{\phi}(k)(y_k - \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta}) = E\{\phi(k)(y_k - \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta})\} + v_k$$

此处 $v_k = \boldsymbol{\phi}(k)(y_k - \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta}) - E\{\boldsymbol{\phi}(k)(y_k - \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta})\}$，显然有 $E\{v_k\} = 0$.

应用随机逼近算法的熟知形式，我们可以得出如下的随机逼近递推辨识算法

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \gamma_k \boldsymbol{\phi}(k+1)[y_{k+1} - \boldsymbol{\phi}(k+1)^\tau \hat{\boldsymbol{\theta}}_k] \quad (2.62)$$

可以证明，这种算法在适当的条件下具有较好的收敛性.

§ 2.7 常见的递推算法所对应的微分方程

在下面的两节中，我们将对常见的递推算法收敛性分析的微分方程方法做一个简单的介绍。

所谓常见的递推算法是指：递推最小二乘法（RLS）、递推辅助变量法（RIV）、递推广义最小二乘法（RGLS）、递推推广的最小二乘法（RELS）和递推极大似然法（RML）。本节的目的是要导出它们所对应的常微分方程。为此，考虑如下的一般模型

$$A(q^{-1})y_k = B(q^{-1})u_k + H(q^{-1})e_k \tag{2.63}$$

其中 $\{e_k\}$ 是零均值的白噪声，而

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \cdots + b_mq^{-m}$$

对于不同的算法，滤波器 $H(q^{-1})$ 有不同的取法，例如：

对于普通的最小二乘法的情形

$$H(q^{-1}) = 1$$

对于广义递推最小二乘法的情形

$$H(q^{-1}) = 1/C(q^{-1})$$

其中

$$C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_pq^{-p}$$

对于推广的递推最小二乘法和递推极大似然法的情形

$$H(q^{-1}) = C(q^{-1})$$

在极大似然法的情形还需假定 e_k 服从正态分布． 在递推辅助变量法的情形，$H(q^{-1})$ 的形式可以不必明确指出．

让我们来详细考虑

$$H(q^{-1}) = C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_pq^{-p}$$

的情形．置

$$\theta^{\tau} = (a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, \cdots, c_p)$$

$$\phi(k)^T = (-y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}, e_{k-1}, \cdots, e_{k-p})$$

则(2.63)式可以写成

$$y_k = \phi(k)^\tau \theta + e_k \tag{2.64}$$

因为 $e_{k-1}, e_{k-2}, \cdots, e_{k-p}$ 是不能直接观测的,所以考虑用残差

$$\varepsilon_{k-i} = y_{k-i} - \hat{\phi}(k-i)^\tau \hat{\theta}(k-i) \quad i = 1, 2, \cdots, p$$

来代替它们,其中 $\hat{\theta}(k)$ 表示 θ 在时刻 k 的估值. 而

$$\hat{\phi}(k) = (-y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}, \varepsilon_{k-1}, \cdots, \varepsilon_{k-p})^\tau$$

显然,这里的 ε_k 和 $\hat{\phi}(k)$ 之间可以交互的被确定. 详细地说,有了 $\varepsilon_{k-1}, \varepsilon_{k-2}, \cdots, \varepsilon_{k-p}$,就可以确定 $\hat{\phi}(k)$; 而有了 $\hat{\phi}(k)$ 以后,通过 $\hat{\theta}(k-1)$,又可确定 ε_k,进而又可确定 $\hat{\phi}(k+1)$,如此等等. 于是可以得出关于 $\hat{\theta}(k)$ 的推广的最小二乘递推算法

$$\hat{\theta}(k) = \hat{\theta}(k-1) + M(k)\varepsilon(k)$$

$$\varepsilon(k) = y_k - \hat{\phi}(k)^\tau \hat{\theta}(k-1)$$

$$M(k) = \frac{P(k-1)\hat{\phi}(k)}{1 + \hat{\phi}(k)^\tau P(k-1)\hat{\phi}(k)}$$

$$P(k) = [I - M(k)\hat{\phi}(k)^\tau]P(k-1)$$

这组递推算法的一组等价的形式是

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{1}{k} R^{-1}(k)\hat{\phi}(k)[y_k - \hat{\phi}(k)^\tau \hat{\theta}(k-1)] \tag{2.65}$$

$$R(k) = R(k-1) + \frac{1}{k}[\hat{\phi}(k)\hat{\phi}(k)^\tau - R(k-1)] \tag{2.66}$$

其中

$$R(k) = \frac{1}{k}\sum_{i=1}^{k} \hat{\phi}(i)\hat{\phi}(i)^\tau$$

事实上,我们有

$$R(k) = \frac{1}{k}\sum_{i=1}^{k} \hat{\phi}(i)\hat{\phi}(i)^\tau$$

$$= \frac{1}{k}\sum_{i=1}^{k-1} \hat{\phi}(i)\hat{\phi}(i)^\tau + \frac{1}{k}\hat{\phi}(k)\hat{\phi}(k)^\tau$$

$$= \frac{k-1}{k} \cdot \frac{1}{k-1} \sum_{i=1}^{k-1} \hat{\boldsymbol{\phi}}(i)\hat{\boldsymbol{\phi}}(i)^{\tau}$$

$$+ \frac{1}{k} \hat{\boldsymbol{\phi}}(k)\hat{\boldsymbol{\phi}}(k)^{\tau}$$

$$= \left(1 - \frac{1}{k}\right) R(k-1) + \frac{1}{k} \hat{\boldsymbol{\phi}}(k)\hat{\boldsymbol{\phi}}(k)^{\tau}$$

$$= R(k-1) + \frac{1}{k} [\hat{\boldsymbol{\phi}}(k)\hat{\boldsymbol{\phi}}(k)^{\tau} - R(k-1)]$$

这就是公式 (2.66)。

注意到模型 (2.64) 中的参量 $\boldsymbol{\theta}$ 使准则

$$J_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{k=1}^{N} [y_k - \hat{\boldsymbol{\phi}}(k)^{\tau}\boldsymbol{\theta}]^2$$

最小化的估值是

$$\hat{\boldsymbol{\theta}}_N = \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\phi}}(k)\hat{\boldsymbol{\phi}}(k)^{\tau}\right]^{-1} \cdot \frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\phi}}(k) y_k \qquad (2.67)$$

于是

$$\hat{\boldsymbol{\theta}}(k) = R(k)^{-1} \cdot \frac{1}{k} \sum_{i=1}^{k} \hat{\boldsymbol{\phi}}(i) y_i$$

由此可见,为证明 (2.65) 式,只需证

$$\frac{1}{k} \sum_{i=1}^{k} \hat{\boldsymbol{\phi}}(i) y_i = R(k)\hat{\boldsymbol{\theta}}(k-1) + \frac{1}{k} \hat{\boldsymbol{\phi}}(k)\{y_k - \hat{\boldsymbol{\phi}}(k)^{\tau}\hat{\boldsymbol{\theta}}(k-1)\}$$

事实上

$$R(k)\hat{\boldsymbol{\theta}}(k-1) + \frac{1}{k} \hat{\boldsymbol{\phi}}(k) y_k - \frac{1}{k} \hat{\boldsymbol{\phi}}(k)\hat{\boldsymbol{\phi}}(k)^{\tau}\hat{\boldsymbol{\theta}}(k-1)$$

$$= \left[\frac{1}{k} \sum_{i=1}^{k} \hat{\boldsymbol{\phi}}(i)\hat{\boldsymbol{\phi}}(i)^{\tau} - \frac{1}{k} \hat{\boldsymbol{\phi}}(k)\hat{\boldsymbol{\phi}}(k)^{\tau}\right] \hat{\boldsymbol{\theta}}(k-1)$$

$$+ \frac{1}{k} \hat{\boldsymbol{\phi}}(k) y_k$$

$$= \frac{1}{k} \sum_{i=1}^{k-1} \hat{\phi}(i)\hat{\phi}(i)^\tau \hat{\theta}(k-1) + \frac{1}{k}\hat{\phi}(k)y_k$$

$$= \frac{k-1}{k} \cdot \frac{1}{k-1} \sum_{i=1}^{k-1} \hat{\phi}(i)\hat{\phi}(i)^\tau \hat{\theta}(k-1) + \frac{1}{k}\hat{\phi}(k)y_k$$

$$= \frac{k-1}{k} R(k-1)\left[\frac{1}{k-1} R(k-1)^{-1} \sum_{i=1}^{k-1} \hat{\phi}(i)y_i\right] + \frac{1}{k}\hat{\phi}(k)y_k$$

$$= \frac{1}{k} \sum_{i=1}^{k} \hat{\phi}(i)y_i$$

这就证明了 (2.65) 式.

公式 (2.65) 和 (2.66) 对于算法的收敛性分析是非常方便的.

显然, $\hat{\phi}(k)$ 和 $\varepsilon(k) = y_k - \hat{\phi}(k)^\tau \hat{\theta}(k-1)$ 都与估值 $\hat{\theta}(0)$, $\cdots$, $\hat{\theta}(k-2)$, $\hat{\theta}(k-1)$ 有关,即应该有

$$\hat{\phi}(k) = \hat{\phi}[k; \hat{\theta}(k-2), \hat{\theta}(k-3), \cdots, \hat{\theta}(0)]$$

$$\varepsilon(k) = \varepsilon[k; \hat{\theta}(k-1), \hat{\theta}(k-2), \cdots, \hat{\theta}(0)]$$

如果把上述表达式中的 $\hat{\theta}(0)$, $\cdots$ $\hat{\theta}(k-2)$, $\hat{\theta}(k-1)$ 都换成 θ, 则 $\hat{\phi}(k)$ 和 $\varepsilon(k)$ 的值都将要发生改变. 所以把它们改记为 $\bar{\hat{\phi}}(k, \theta)$ 和 $\varepsilon(k, \theta)$, 即有

$$\bar{\hat{\phi}}(k, \theta) = [-y_{k-1}, \cdots, -y_{k-n}, u_k, u_{k-1}, \cdots, u_{k-m}, \varepsilon(k-1, \theta), \cdots, \varepsilon(k-p, \theta)]^\tau$$

$$\varepsilon(k, \theta) = y_k - \bar{\hat{\phi}}(k, \theta)^\tau \theta$$

假定 $\{u_k\}$ 是平稳序列, $\{e_k\}$ 是零均值的白噪声, 于是 $\{\varepsilon(k, \theta)\}$ 和 $\{\bar{\hat{\phi}}(k, \theta)\}$ 都是平稳序列. 置

$$f(\theta) = E\{\bar{\hat{\phi}}(k, \theta)\varepsilon(k, \theta)\}$$

$$G(\theta) = E\{\bar{\hat{\phi}}(k, \theta)\bar{\hat{\phi}}(k, \theta)^\tau\}$$

于是可以把 (2.65) 式写成

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{1}{k} R^{-1}(k)\hat{\phi}(k)\varepsilon(k)$$

$$= \hat{\theta}(k-1) + \frac{1}{k} R^{-1}(k) f[\hat{\theta}(k-1)]$$

$$+ \frac{1}{k} R^{-1}(k)\{\hat{\phi}(k)\varepsilon(k) - \bar{\hat{\phi}}(k, \hat{\theta}(k-1))$$

$$\cdot \varepsilon(k, \hat{\theta}(k-1))\} + \frac{1}{k} R^{-1}(k)$$

$$\cdot \{\bar{\hat{\phi}}(k, \hat{\theta}(k-1))\varepsilon(k, \hat{\theta}(k-1))$$

$$- f[\hat{\theta}(k-1)]\}$$

令

$$\beta_{k-1} = R^{-1}(k)\{\hat{\phi}(k)\varepsilon(k) - \bar{\hat{\phi}}(k, \hat{\theta}(k-1))\varepsilon(k, \hat{\theta}(k-1))\}$$

$$\xi_{k-1} = R^{-1}(k)\{\bar{\hat{\phi}}(k, \hat{\theta}(k-1))\varepsilon(k, \hat{\theta}(k-1)) - f[\hat{\theta}(k-1)]\}.$$

则有

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{1}{k} R^{-1}(k) f[\hat{\theta}(k-1)]$$

$$+ \frac{1}{k}\beta_{k-1} + \frac{1}{k}\xi_{k-1}$$

注意到当 $k \to \infty$ 时，在适当的条件下由 (2.65) 式可见 $\hat{\theta}(k) \approx \hat{\theta}(k-1)$，故可以得出

$\{\beta_k\}$ 以概率 1 有界,并且

$$\lim_{k\to\infty} \beta_k = 0 \qquad \text{a.s}$$

此外,由于

$$E\{\xi_k | \hat{\theta}(0), \cdots, \hat{\theta}(k-1)\} = 0$$

并且可能有

$$E\{\xi_k | \xi_0, \xi_1, \cdots, \xi_{k-1}\} = E\{\xi_k | \hat{\theta}(0), \cdots, \hat{\theta}(k-1)\}$$

所以可以说明

$$\lim_{k\to\infty} \xi_k = 0 \qquad \text{a.s}$$

总之,我们能够得出算法 (2.65) 所对应的微分方程是

$$\frac{d\theta(t)}{dt} = R^{-1}(t) f[\theta(t)]$$

类似地可以说明,算法 (2.66) 所对应的微分方程是

$$\frac{dR(t)}{dt} = G[\hat{\theta}(t)] - R(t)$$

其实，我们可以把递推最小二乘法、递推极大似然法、递推广义最小二乘法、递推推广最小二乘法、递推辅助变量法等五种递推算法，写成下述的统一形式

$$\hat{\theta}(k+1) = \hat{\theta}(k) + M(k+1)\varepsilon(k+1)$$

$$M(k+1) = \frac{P(k)\boldsymbol{z}(k+1)}{\lambda(k+1) + \boldsymbol{\varphi}(k+1)^{\tau}P(k)\boldsymbol{z}(k+1)}$$

$$P(k+1) = \frac{1}{\lambda(k+1)}[I - M(k+1)\boldsymbol{\varphi}(k+1)^{\tau}]P(k)$$

$$\lambda(k+1) = \lambda_0\lambda(k) + (1-\lambda_0)$$

λ_0 是适当的常数. 对于不同的算法，$\boldsymbol{z}(k)$，$\boldsymbol{\varphi}(k)$ 和 $\boldsymbol{\varepsilon}(k)$ 有不同的取法，详细叙述可从文献[10]中找到.

类似于前面的讨论，可以说明上述形式的递推算法所对应的微分方程是

$$\frac{d\boldsymbol{\theta}(t)}{dt} = R(t)f[\boldsymbol{\theta}(t)] \tag{2.68}$$

$$\frac{dR(t)}{dt} = G[\boldsymbol{\theta}(t)] - R(t) \tag{2.69}$$

其中

$$f(\boldsymbol{\theta}) = E\{\bar{\boldsymbol{z}}(k,\boldsymbol{\theta})\bar{\varepsilon}(k,\boldsymbol{\theta})\}$$

$$G(\boldsymbol{\theta}) = E\{\bar{\boldsymbol{z}}(k,\boldsymbol{\theta})\bar{\varphi}(k,\boldsymbol{\theta})^{\tau}\}$$

而 $\{\bar{\varepsilon}(k,\boldsymbol{\theta})\}$，$\{\bar{\boldsymbol{\varphi}}(k,\boldsymbol{\theta})\}$ 和 $\{\bar{\boldsymbol{z}}(k,\boldsymbol{\theta})\}$ 都是平稳序列. 它们是由相应的递推算法中的

$$\boldsymbol{\varepsilon}(k) = \boldsymbol{\varepsilon}(k; \hat{\boldsymbol{\theta}}(0), \cdots, \hat{\boldsymbol{\theta}}(k-1))$$

$$\boldsymbol{\varphi}(k) = \boldsymbol{\varphi}(k; \hat{\boldsymbol{\theta}}(0), \cdots, \hat{\boldsymbol{\theta}}(k-2))$$

$$\boldsymbol{z}(k) = Z(k; \hat{\boldsymbol{\theta}}(0), \cdots, \hat{\boldsymbol{\theta}}(k-2))$$

将估值 $\hat{\boldsymbol{\theta}}(0), \cdots, \hat{\boldsymbol{\theta}}(k-2), \hat{\boldsymbol{\theta}}(k-1)$ 用 $\boldsymbol{\theta}$ 代替后所得到的.

§2.8 常见的递推算法的收敛性

本节介绍萨德斯特隆姆（T. Söderström）和雍（L. Ljung）

等人利用微分方程对常见的五种递推算法的收敛性进行分析的结果.

定理 2.7 设

$$\frac{d\theta(t)}{dt} = R^{-1}(t)f[\theta(t)] \tag{2.70}$$

$$\frac{dR(t)}{dt} = G[\theta(t)] - R(t) \tag{2.71}$$

是五种常见的递推算法的统一形式所对应的微分方程,其中

$$f(\theta) = E\{\bar{z}(k,\theta)\varepsilon(k,\theta)\}$$
$$G(\theta) = E\{\bar{z}(k,\theta)\bar{\varphi}(k,\theta)^{\tau}\}$$

$\{\varepsilon(k,\theta)\}$, $\{\bar{\varphi}(k,\theta)\}$, $\{\bar{z}(k,\theta)\}$ 皆为平稳序列, 它们的定义如前节所述. 于是我们有:

1) 如果 $(\theta^*, G(\theta^*))$ 是上述微分方程组的大范围渐近稳定平稳点,则

$$\lim_{k\to\infty}\hat{\theta}(k) = \theta^* \qquad \text{a.s}$$

2) 递推算法所有收敛点的集合是集合 D_s 的一个子集. 此处 $D_s = \{\theta: f(\theta) = 0\}$

3) $\hat{\theta}(k)$ 收敛到 θ^* 的必要条件是上述的微分方程在 θ^* 的一个邻域内局部稳定. 即

$$[G(\theta^*)]^{-1}\frac{d}{d\theta}f(\theta)\bigg|_{\theta=\theta^*}$$

的所有特征值皆有非正的实部.

对这一定理,我们仅作如下的简要说明.

关于结论 1), 由于递推算法

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{1}{k}R^{-1}(k)\varphi(k)\varepsilon(k)$$

可以看成是微分方程

$$\frac{d\theta(t)}{dt} = R^{-1}(t)f[\theta(t)]$$

的欧拉 (Euler) 解. 所以上述微分方程的轨线可以解释为递推算

法的渐近路线．而 θ^* 是它的大范围的渐近稳定平稳点，于是对于上述方程的有界渐近稳定解，由渐近稳定性有

$$\lim_{t\to\infty}\theta(t)=\theta^*$$

所以有

$$\lim_{k\to\infty}\hat{\theta}(k)=\theta^*\quad \text{a.s}$$

关于结论 2)．由于

$$R^{-1}(k)\varphi(k)\varepsilon(k)=k\{\hat{\theta}(k)-\hat{\theta}(k-1)\}$$

如果算法是收敛的，则只要 k 充分大就可能有 $k\{\hat{\theta}(k)-\hat{\theta}(k-1)\}\approx 0$，所以在递推算法的收敛点 θ 处，应该有 $f(\theta)=0$．

关于结论 3)．先把微分方程 (2.70) 和 (2.71) 在点 (θ^*, R^*) 处线性化，此处 $R^*=G(\theta^*)$．由泰勒公式我们有

$$\frac{d\theta(t)}{dt}\approx G(\theta^*)^{-1}f(\theta^*)+G(\theta^*)^{-1}\left.\frac{df(\theta)}{d\theta}\right|_{\theta^*}\Delta\theta - R^{*-2}f(\theta^*)\Delta R \tag{2.72}$$

$$\frac{dR(t)}{dt}\approx G(\theta^*)-R^*+\frac{d}{d\theta}G(\theta)|_{\theta^*}\Delta\theta - I\cdot\Delta R \tag{2.73}$$

其中 $\Delta\theta=\theta-\theta^*$，$\Delta R=R-R^*=R-G(\theta^*)$．由于 $f(\theta^*)=0$，所以方程 (2.72) 变成了

$$\frac{d\Delta\theta(t)}{dt}=G(\theta^*)^{-1}\left.\frac{df(\theta)}{d\theta}\right|_{\theta^*}\Delta\theta(t) \tag{2.74}$$

如果有 $\lim\limits_{k\to\infty}\hat{\theta}(k)=\theta^*$，则由于原方程必须在 θ^* 处是局部渐近稳定的，于是方程 (2.74) 在 $\Delta\theta(t)=0$ 处稳定．由微分方程稳定性理论知，若上述的线性微分方程 (2.74) 在 $\Delta\theta=0$ 处是稳定的，必须有

$$G(\theta^*)^{-1}\left.\frac{df(\theta)}{d\theta}\right|_{\theta^*}$$

的所有特征根皆有非正的实部． ■

这个定理完全清楚地描述了常见递推算法的收敛性和相应的

微分方程的稳定性之间的关系．以下我们应用上述定理来分析五种常见的递推算法的收敛性．

对于递推最小二乘法需要指出的是，参量 $\boldsymbol{\theta}$ 的真值 $\boldsymbol{\theta}^*$ 必为 $f(\boldsymbol{\theta})$ 的零点，即 $f(\boldsymbol{\theta}^*)=0$．并且递推算法所对应的微分方程

$$\frac{d\boldsymbol{\theta}(t)}{dt}=R^{-1}(t)f[\boldsymbol{\theta}(t)] \tag{2.75}$$

$$\frac{dR(t)}{dt}=G[\boldsymbol{\theta}(t)]-R(t) \tag{2.76}$$

在 $\boldsymbol{\theta}^*$ 处是大范围渐近稳定的，此处

$$f(\boldsymbol{\theta})=E\{\bar{\varphi}(k,\boldsymbol{\theta})\bar{\varepsilon}(k,\boldsymbol{\theta})\}$$
$$G(\boldsymbol{\theta})=E\{\bar{\varphi}(k,\boldsymbol{\theta})\bar{\varphi}(k,\boldsymbol{\theta})^{\tau}\}$$

事实上，我们有

$$\begin{aligned}f(\boldsymbol{\theta}^*)&=E\{\bar{\varphi}(k,\boldsymbol{\theta}^*)\bar{\varepsilon}(k,\boldsymbol{\theta}^*)\}\\&=E\{\bar{\varphi}(k,\boldsymbol{\theta}^*)[y_k-\bar{\varphi}(k,\boldsymbol{\theta}^*)^{\tau}\boldsymbol{\theta}^*]\}\\&=E\{\bar{\varphi}(k,\boldsymbol{\theta}^*)e_k\}+E\{\bar{\varphi}(k,\boldsymbol{\theta}^*)\\&\quad\cdot\bar{\varphi}(k,\boldsymbol{\theta}^*)^{\tau}[\boldsymbol{\theta}^*-\boldsymbol{\theta}]\}\end{aligned}$$

由于 e_k 与 $\bar{\varphi}(k,\boldsymbol{\theta}^*)$ 独立，并且 $E\{e_k\}=0$，所以

$$f(\boldsymbol{\theta}^*)=0$$

为说明方程 (2.75) 在 $\boldsymbol{\theta}^*$ 处的渐近稳定性，引入如下的李亚普诺夫 (Lyapunov) 函数

$$V(\boldsymbol{\theta},R)=\frac{1}{2}E\{[\bar{\varepsilon}(k,\boldsymbol{\theta})-e_k]^2\}$$

沿着方程 (2.75) 和 (2.76) 的解轨线，我们有

$$\begin{aligned}\frac{d}{dt}V(\boldsymbol{\theta},R)&=E\left\{[\bar{\varepsilon}(k,\theta)-e_k]\frac{d}{d\theta}\bar{\varepsilon}(k,\boldsymbol{\theta})\right\}\frac{d\boldsymbol{\theta}}{dt}\\&=E\left\{\bar{\varepsilon}(k,\boldsymbol{\theta})\frac{d}{d\boldsymbol{\theta}}\bar{\varepsilon}(k,\boldsymbol{\theta})\right\}R^{-1}f(\boldsymbol{\theta})\end{aligned}$$

而在最小二乘法的情形，$\bar{\varphi}(k)$ 不依赖于参量 $\boldsymbol{\theta}$．故有

$$\frac{d}{d\boldsymbol{\theta}}\bar{\varphi}(k,\boldsymbol{\theta})=0$$

所以

$$\frac{d}{d\theta}\bar{\varepsilon}(k,\boldsymbol{\theta}) = -\bar{\varphi}(k,\boldsymbol{\theta})$$

从而有

$$\frac{d}{dt}V(\boldsymbol{\theta},R) = -f(\boldsymbol{\theta})^{\tau}R^{-1}f(\boldsymbol{\theta}) < 0$$

即方程(2.75)在$\boldsymbol{\theta}^*$处是大范围渐近稳定的. 所以

$$\lim_{k\to\infty}\hat{\boldsymbol{\theta}}(k) = \boldsymbol{\theta}^* \quad \text{a.s}$$

对于递推辅助变量法，由于辅助变量的选取必须满足辅助变量法所提出的条件，这就保证了离线的辅助变量法具有一致性.而递推的辅助变量法与它的离线算法是等价的，所以它必然收敛.

对于其余的三种算法，由于有

$$\bar{\varepsilon}(k,\theta^*) = e_k$$

所以只要e_k与$\bar{Z}(k,\boldsymbol{\theta}^*)$独立，就有

$$\begin{aligned} f(\boldsymbol{\theta}^*) &= E\{\bar{Z}(k,\boldsymbol{\theta}^*)\bar{\varepsilon}(k,\boldsymbol{\theta}^*)\} \\ &= E\{\bar{Z}(k,\boldsymbol{\theta}^*)\}E\{e_k\} = 0 \end{aligned}$$

关于相应的微分方程的稳定性，以极大似然法为例. 由于在递推算法的表达式中有

$$\hat{\boldsymbol{\theta}} = [\hat{a}_1,\cdots,\hat{a}_n,\hat{b}_0,\hat{b}_0,\cdots,\hat{b}_m,\hat{C}_1,\cdots\hat{C}_p]^{\tau}$$

$$\begin{aligned} \boldsymbol{z}(k) = \boldsymbol{\varphi}(k) = [&-y_{k-1},\cdots,-y_{k-n},u_k,u_{k-1}, \\ &\cdots,u_{k-m},\varepsilon(k-1),\cdots,\varepsilon(k-P)]^{\tau}\frac{1}{C(q^{-1})} \end{aligned}$$

而预报误差$\varepsilon(k)$由下式给出

$$\varepsilon(k) = \frac{1}{\hat{C}(q^{-1})}[\hat{A}(q^{-1})y_k - \hat{B}(q^{-1})u_k]$$

于是只要取李亚普诺夫函数为

$$V(Q,R) = \frac{1}{2}E\{[\bar{\varepsilon}(k,\theta) - e(k)]^2\}$$

注意到

$$\frac{\partial\varepsilon(k)}{\partial\hat{a}_i} = \frac{1}{\hat{C}(q^{-1})}y_{k-i} \qquad \frac{\partial\varepsilon(k)}{\partial\hat{b}_i} = -\frac{1}{\hat{C}(q^{-1})}u_{k-i}$$

$$\frac{\partial \varepsilon(k)}{\partial \hat{C}_h} = \frac{-q^{-h}}{\hat{C}(q^{-1})^2}[\hat{A}(q^{-1})y_k - \hat{B}(q^{-1})u_k]$$

$$= -\frac{1}{\hat{C}(q^{-1})}\varepsilon(k-h)$$

故有

$$\frac{d}{d\theta}\varepsilon(k,\theta)^{\tau} = -\bar{Z}(k,\theta)$$

从而

$$\frac{d}{dt}V(\theta,k) = -f(\theta)^{\tau}R^{-1}f(\theta) < 0$$

这就得出了递推极大似然法所对应的微分方程在 θ^* 处的大范围渐近稳定性，从而有

$$\lim_{k\to\infty}\hat{\theta}(k) = \theta^* \qquad \text{a.s}$$

类似地可以对递推广义最小二乘法和递推推广最小二乘法的收敛性进行分析.

§2.9 模型定阶的 AIC 准则

模型结构的确定是系统辨识的一个重要组成部分. 对一般系统而言，尚没有一个统一且有效的确定结构的方法. 然而对于线性非时变状态空间模型和 ARMA 类型的模型，情况就不同了. 关于状态空间模型结构确定的理论，在这里就不介绍了.

ARMA 模型的结构完全被它的阶数所确定，确定这类模型的阶已有一系列有效的方法，赤池 (Akaike) 所提出的 AIC 准则 (最小信息准则)可以作为这类方法的一个代表. 本节我们重点介绍这一准则.

先引入库尔贝克-莱布勒 (Kullback-Leibler) 信息测度的概念.

设 $p(z)$ 和 $q(z)$ 是两个概率密度函数，我们称

$$I(p,q) = -\int\left\{\ln\frac{q(z)}{p(z)}\right\}p(z)dz \tag{2.77}$$

为关于 $p(z)$ 和 $q(z)$ 的平均信息，或库尔贝克-莱布勒信息测度.

结论 库尔贝克-莱布勒信息测度恒满足

$$I(p, q) \geqslant 0 \tag{2.78}$$

而 $I(p, q) = 0$ 的充分必要条件是 $q(z) \equiv p(z)$.

让我们应用信息测度这一工具导出关于模型阶的估计的 AIC 准则,即最小信息准则. 设模型中含有待定参数 $\boldsymbol{\theta}$, 它的维数直接描写了模型的阶数, $\boldsymbol{\Theta}$ 是 $\boldsymbol{\theta}$ 的变化范围 ($\boldsymbol{\theta}$ 的维数和值都可能变化).

z 是模型的输出,设输出数据是

$$Z_N = \{z_1, z_2, \cdots, z_N\}$$

$p(Z_N|\boldsymbol{\theta})$ 是条件分布密度. $\boldsymbol{\theta}_0$ 是 $\boldsymbol{\theta}$ 的真值 (包括真实的维数), $\boldsymbol{\theta}$ 的维数反映了模型的真实阶数. 如果 $\boldsymbol{\theta}_0$ 已知, 则与真实模型相对应的条件密度是 $p(Z_N|\boldsymbol{\theta}_0)$. 与密度函数 $p(Z_N|\boldsymbol{\theta})$ 和 $p(Z_N|\boldsymbol{\theta}_0)$相应的库尔贝克-莱布勒信息测度是

$$I_N(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = -\int \ln \left\{\frac{p(Z_N|\boldsymbol{\theta})}{p(Z_N|\boldsymbol{\theta}_0)}\right\} p(Z_N|\boldsymbol{\theta}_0) dZ_N \tag{2.79}$$

就参数估计的意义来说,我们自然希望 $\boldsymbol{\theta}$ 的估值 $\hat{\boldsymbol{\theta}}$ 满足

$$I_N(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}) = \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} I_N(\boldsymbol{\theta}_0, \boldsymbol{\theta})$$

如果置

$$S(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \int \{\ln p(Z_N|\boldsymbol{\theta})\} p(Z_N|\boldsymbol{\theta}_0) dZ_N$$

则由 (2.70) 式有

$$I_N(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = S(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) - S(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \tag{2.80}$$

由此可见, $\hat{\boldsymbol{\theta}}$ 使 $I_N(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}})$ 最小等价于使 $S(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}})$ 最大, 也就是使 $\ln p(Z_N|\boldsymbol{\theta})$ 最大. 所以 $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(Z_N)$ 是极大似然估计.

可以看出, $\boldsymbol{\theta}_0$ 使

$$I_N(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 0$$

即 $\boldsymbol{\theta}_0$ 是 $I_N(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ 的最小值点,从而有

$$\frac{\partial}{\partial \boldsymbol{\theta}} I_N(\boldsymbol{\theta}_0, \boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}_0} = 0$$

故有

$$I_N(\theta_0, \theta_0 + \Delta\hat{\theta}) = \frac{1}{2} \Delta\hat{\theta}^\tau J_N(\hat{\theta}_0)\Delta\hat{\theta} + o(\|\Delta\hat{\theta}\|^2) \qquad (2.81)$$

其中

$$J_N(\theta_0) = -\int \frac{\partial}{\partial\theta} \left\{\frac{\partial}{\partial\theta} \ln p(Z_N|\theta)\right\}\Bigg|_{\theta_0} p(Z_N|\theta_0)dZ_N$$

$$= \int\left[\frac{\partial}{\partial\theta} \ln p(Z_N|\theta) \frac{\partial}{\partial\theta} \ln p(Z_N|\theta)^\tau\right]\Bigg|_{\theta_0}$$

$$\cdot p(Z_N|\theta_0)dZ_N$$

即 $J_N(\theta_0)$ 是费歇耳 (Fisher) 信息矩阵，$\Delta\theta = \theta - \theta_0$. 如果 θ 的维数是 k，则当样本充分大时，$\Delta\theta^\tau J_N(\theta_0)\Delta\theta$ 渐近于 χ_k^2 分布.

注意到 $o(\|\Delta\theta\|^2) \to 0$，所以由 (2.81) 式得出

$$2I_N(\theta_0, \hat{\theta}) \approx (\hat{\theta} - \theta_0)^\tau J_N(\theta_0)(\hat{\theta} - \theta_0) \sim \chi_k^2$$

如果用 E_∞ 表示对样本极限分布所取的均值，则有

$$2E_\infty\{I_N(\theta_0, \hat{\theta})\} = k$$

但

$$E_\infty\{I_N(\theta_0, \hat{\theta})\} = E_\infty\left\{\int[\ln p(z_N|\theta_0)]p(Z_N|\theta_0)dZ_N\right\}$$

$$- E_\infty\left\{\int [\ln p(Z_N|\hat{\theta})]\, p(Z_N(\theta_0)dZ_N\right\}$$

故有

$$-2E_\infty\left\{\int [\ln p(Z_N|\hat{\theta})]p(Z_N|\theta_0)dZ_N\right\}$$

$$= 2E_\infty\{I_N(\theta_0,\hat{\theta})\} - 2E_\infty\left\{\int [\ln p(Z_N|\theta_0)]\right.$$

$$\left.\cdot\, p(Z_N|\theta_0)dZ_N\right\}$$

$$= k - 2E_\infty\left\{\int [\ln p(Z_N|\theta_0)]p(Z_N|\theta_0)dZ_N\right\} \qquad (2.82)$$

现在对函数 $\ln L(\theta) = \ln p(Z_N|\theta)$ 在 $\hat{\theta}$ 处进行泰勒展开，有

$$\ln L(\theta) = \ln L(\hat{\theta}) + \left(\frac{\partial}{\partial\theta} \ln L(\theta)\Bigg|_{\hat{\theta}}^\tau\right)(\theta - \hat{\theta})$$

$$+ \frac{1}{2} (\theta - \hat{\theta})^\tau \frac{\partial^2}{\partial\theta\partial\theta^\tau} \ln L(\theta)\Bigg|_{\hat{\theta}} \cdot (\theta - \hat{\theta})$$

$$+ o(\|\Delta\boldsymbol{\theta}\|^2)$$

注意到 $\hat{\boldsymbol{\theta}}$ 是极大似然估计，所以有

$$\frac{\partial}{\partial\boldsymbol{\theta}} \ln L(\boldsymbol{\theta})\Big|_{\hat{\theta}} = 0$$

故有

$$\ln L(\boldsymbol{\theta}_0) \approx \ln L(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})^{\tau} Q(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})$$

其中

$$Q(\hat{\boldsymbol{\theta}}) = \frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{\partial}{\partial\boldsymbol{\theta}} \ln L(\boldsymbol{\theta})\right)\Big|_{\hat{\theta}}$$

于是

$$J_N(\hat{\boldsymbol{\theta}}) = -E\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{\partial}{\partial\boldsymbol{\theta}} \ln L(Z_N, \boldsymbol{\theta})\right)\right\}\Big|_{\hat{\theta}} = -E\{Q(\hat{\boldsymbol{\theta}})\}$$

从而

$$\begin{aligned} & 2\int \{\ln P(Z_N|\hat{\boldsymbol{\theta}})\} P(Z_N|\boldsymbol{\theta}_0) dZ_N \\ & - 2\int [\ln p(Z_N|\boldsymbol{\theta}_0)] P(Z_N|\boldsymbol{\theta}_0) dZ_N \\ & \approx (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^{\tau} J_N(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim \chi_k^2 \end{aligned}$$

由此不难看出，选取 $\hat{\boldsymbol{\theta}}$ 使 (2.82) 式的左端最小等价于，选取 $\hat{\boldsymbol{\theta}}$ 使

$$k - 2E_{\infty}\left\{\int \ln P(Z_N|\boldsymbol{\theta}_0) P(Z_N|\boldsymbol{\theta}_0) dZ_N\right\} \tag{2.83}$$

最小，其中

$$\begin{aligned} & 2E_{\infty}\left\{\int [\ln p(Z_N|\boldsymbol{\theta}_0)] p(Z_N|\boldsymbol{\theta}_0) dZ_N\right\} \\ & \quad \approx E_{\infty}\left\{2\int [\ln p(Z_N|\hat{\boldsymbol{\theta}})] p(Z_N|\boldsymbol{\theta}_0) dZ_N\Big|_{\theta_0}\right\} - k \end{aligned}$$

这说明选取 $\hat{\boldsymbol{\theta}}$ 使 (2.83) 式最小相当于使

$$-2E\theta_0\left\{\int[\ln p(Z_N|\theta)]p(Z_N|\theta_0)dZ_N\right\}+2k$$

取最小. 由此可见,只需使

$$-2\ln p(Z_N|\hat{\theta})+2k$$

取最小. 故我们定义 AIC 准则为

$$\mathrm{AIC}(k)=-2\ln L(Z_N,\hat{\theta})+2k$$

使 AIC(k) 为极小的 k_0,即作为模型的阶数的估值.

§ 2.10 模型定阶的 FPE 准则

我们引出 AR 模型

$$y_k+a_1y_{k-1}+\cdots+a_ny_{k-n}=e_k \tag{2.84}$$

定阶的最终预报误差 (FPE) 准则, 其中模型的阶数 n 是待估计的. 不妨设 y_k 的维数是 1,置

$$Y_N=\{y_1,y_2,\cdots,y_N\}$$

N是样本长度,用 σ_e^2 表示 e_k 的方差. 由于 e_k 是零均值的白噪声,所以有 $E\{e_k\}=0$, 故有 $\sigma_e^2=E\{e_k^2\}$.

设 $\theta=(a_1,a_2,\cdots,a_n)^\tau$,于是

$$P(Y_N|\theta)=\prod_{k=1}^{N}P(y_k|Y_{k-1},\theta)$$

当 e_k 是高斯白噪声时,可得

$$P(y_k|Y_{k-1},\theta)=\frac{1}{\sigma_e\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{e_k}{\sigma_e}\right)^2\right\}$$

从而

$$p(Y_N|\theta)=\frac{1}{\sigma_e^N(2\pi)^{\frac{N}{2}}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{N}e_k^2\sigma_e^{-2}\right\}$$

取对数得

$$\ln p(Y_N|\theta)=-\frac{N}{2}(\ln\sigma_e^2+\ln 2\pi)-\frac{1}{2}\sum_{k=1}^{N}e_k^2\sigma_k^{-2} \tag{2.85}$$

因为 σ_e^2 也可能是未知的，所以把似然函数 $p(Y_N|\boldsymbol{\theta})$ 改写成 $L(\boldsymbol{\theta}, \sigma_e^2)$. 于是有

$$-\ln L(\boldsymbol{\theta}, \sigma_e^2) = \frac{N}{2}(\ln\sigma_e^2 + \ln 2\pi) + \frac{1}{2}\sigma_e^{-2}\sum_{k=1}^{N} e_k^2 \quad (2.86)$$

对于固定的 $\boldsymbol{\theta}$，对上式两端求关于 σ_e^2 的偏导数得

$$-\frac{\partial \ln L(\boldsymbol{\theta}, \sigma_e^2)}{\partial \sigma_e^2} = \frac{N}{2}\sigma_e^{-2} - \frac{1}{2}\sigma_e^{-4}\sum_{k=1}^{N} e_k^2$$

于是可得 σ_e^2 的极大似然估值

$$\hat{\sigma}_e^2 = R^N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{k=1}^{N} e_k^2 \quad (2.87)$$

把它代入 (2.86) 式得

$$-\ln \max_{\sigma_e^2} L(\boldsymbol{\theta}, \sigma_e^2) = \frac{N}{2}\ln R^N(\boldsymbol{\theta}) + \frac{N}{2}(1 + \ln 2\pi)$$

设 $\hat{\boldsymbol{\theta}}_N$ 是 $\boldsymbol{\theta}$ 的极大似然估值，由上式可见，这样的 $\hat{\boldsymbol{\theta}}_N$ 必满足

$$R^N(\hat{\boldsymbol{\theta}}_N) = \min_{\boldsymbol{\theta}} R^N(\boldsymbol{\theta})$$

假定

$$\lim_{N\to\infty} \hat{\boldsymbol{\theta}}_N = \hat{\boldsymbol{\theta}}_\infty = \boldsymbol{\theta}^* \qquad \text{a.s}$$

其中 $\boldsymbol{\theta}^*$ 表示参量 $\boldsymbol{\theta}$ 的真值(在§2.8 中，我们已给出了递推极大似然法强一致的充分条件).

置

$$\boldsymbol{\phi}(k) = (-y_{k-1}, -y_{k-2}, \cdots, -y_{k-n})^\tau$$

则可把模型 (2.84) 写成下述形式

$$y_k = \boldsymbol{\phi}(k)^\tau\boldsymbol{\theta} + e_k$$

令

$$\hat{y}_N(k) = \boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}_N$$

称 $\hat{y}_N(k)$ 为 y_k 依据数据 Y_N 的预报值(此时假定 $K > N$). 则

$$\hat{y}_N(k) - Y_k$$

是预报误差. 置

$$ER(\hat{\theta}_N) = E(y_k - \hat{y}_N(k))^2$$

即它是预报误差平方的期望值.

把 $R(\hat{\theta}_N)$ 在 θ^* 处按泰勒公式展开,有

$$\begin{aligned} R(\hat{\theta}_N) = {} & R(\theta^*) + \nabla R(\theta^*)^\tau(\hat{\theta}_N - \theta^*) \\ & + \frac{1}{2}(\hat{\theta}_N - \theta^*)^\tau \nabla^2 R(\theta^*)(\hat{\theta}_N - \theta^*) \\ & + o(\|\hat{\theta}_N - \theta^*\|^2) \end{aligned} \tag{2.88}$$

其中 $\nabla R(\theta)$ 表示 $R(\theta)$ 关于 θ 的梯度, $\nabla^2 R(\theta)$ 表示 $R(\theta)$ 关于 θ 的二阶偏导数矩阵.

置

$$Q(\hat{\theta}_N, \theta^*) = \frac{1}{2}(\hat{\theta}_N - \theta^*)^\tau \nabla^2 R(\theta^*)(\hat{\theta}_N - \theta^*)$$

略去 (2.88) 式中的高阶项,有

$$R(\hat{\theta}_N) \approx R(\theta^*) + Q(\hat{\theta}_N, \theta^*) \tag{2.89}$$

进一步的分析可以看出,随机变量

$$\frac{NQ(\hat{\theta}_N, \theta^*)}{R(\theta^*)}$$

的分布,当 $N \to \infty$ 时,以自由度为 n (即模型中包含的独立参数的数目)的 χ^2-分布为极限. 于是对 (2.89) 式两端取数学期望得

$$E\{R(\hat{\theta}_N)\} \approx R(\theta^*)\left(1 + \frac{n}{N}\right) \tag{2.90}$$

即我们得到了预报误差平方平均值的一个近似表达式, 但 $R(\theta^*)$ 的一个无偏的估值是

$$\frac{N}{N-n} R^N(\hat{\theta}_N)$$

用这个量代替 (2.90) 式中的 $R(\theta^*)$,于是得出

$$E\{R(\hat{\theta}_N)\} \approx R^N(\hat{\theta}_N)\frac{N+n}{N-n} = R^N(\hat{\theta}_N)\frac{1+\dfrac{n}{N}}{1-\dfrac{n}{N}} \tag{2.91}$$

我们称 (2.91) 式的右端的表达式为最终预报误差的估值,并把它

记成 FPE.

对于多输入多输出系统也可以导出关于它的 FPE 的与(2.91)式相同的表达式.

利用最终预报误差来确定 AR 模型 (2.84) 的阶数 n，就是要寻求这样的 n，它使得

$$\mathrm{FPE}(n) = \min_m R^N(\hat{\boldsymbol{\theta}}_N) \frac{1 + \frac{m}{N}}{1 - \frac{m}{N}}$$

注意到

$$\left(1 + \frac{n}{N}\right)\left(1 - \frac{n}{N}\right)^{-1} \approx 1 + \frac{2n}{N}$$

所以最终预报误差 FPE(n) 还可以演化出另一种形式

$$\mathrm{FPE}_2(n) = \left(1 + \frac{2n}{N}\right) R^N(\hat{\boldsymbol{\theta}}_N)$$

更一般地有

$$\mathrm{FPE}_\alpha(n) = \left(1 + \frac{\alpha n}{N}\right) R^N(\hat{\boldsymbol{\theta}}_N)$$

α 为实数.

这里的 FPE 准则和前节所介绍 AIC 准则都避免了统计检验中由于选取置信度而产生的人为性，为时间序列模型的定阶带来了许多方便. 然而，它们仍存在某些不足，以 AIC 准则为例，有人已证明了由它所给出的阶数的估计不具有一致性. 即当样本长度 N 趋于无穷时，用 AIC 准则确定的模型阶数估值，不能按概率收敛到真值. 为了对 AIC 准则进行改进，又提出了 BIC 准则

$$\mathrm{BIC}(K) = N \log \hat{\sigma}_\varepsilon^2 + K \log N$$

其中 $K = n + m + 1$，n 是模型的自回归阶数，m 是滑动平均阶数，N 是样本长度，$\hat{\sigma}_\varepsilon^2$ 是 σ_ε^2 的估值.

BIC 准则已使其估值在一定条件下，具有一致性了.

第三章　解决预报问题的古典方法

本章，我们介绍从本世纪 40 年代开始发展起来的一些预报理论和方法.

§ 3.1　时间序列的预报问题

设 $\{x_k\}$ 是一个时间序列，而资料

$$x_k, x_{k-1}, \cdots, x_{k-n}$$

是已知的，其中 n 可以是有限数，也可以是无限数. 我们把依据资料 $\{x_k, x_{k-1}, \cdots, x_{k-n}\}$ 对 x_{k+m} 的值进行估计的问题称为预报问题，此处 m 是正整数. 所得到的值称为 x_{k+m} 的预报估值，简称为预报值，记作 $\hat{x}(k+m|k)$ 或简记作 $\hat{x}_{k+m}$. 既然 $\hat{x}_{k+m}$ 的确定主要是依据资料集 $\{x_k, x_{k-1}, \cdots, x_{k-n}\}$ 所提供的信息，所以它应是变量 $x_k, x_{k-1}, \cdots, x_{k-n}$ 的某个函数

$$\hat{x}_{k+m} = F(x_k, x_{k-1}, \cdots, x_{k-n})$$

于是，寻求预报公式的问题就变成了确定函数 $F(\cdot)$ 的具体表达式的问题.

1. 预报准则

怎样选取 $F(\cdot)$ 才是合理的呢？为回答此问题，必须首先确定一种准则. 依据这种准则所确定的 $F(\cdot)$，应该使预报误差

$$\tilde{x}_{k+m} = x_{k+m} - \hat{x}_{k+m}$$

的某种函数尽可能地小. 由于 x_{k+m} 和 $\hat{x}_{k+m}$ 都是随机的量，所以，我们自然地把这种准则取为使预报误差的平方平均值为最小. 即令

$$\mathscr{B}(k, m) = E\{|x_{k+m} - \hat{x}_{k+m}|^2\} \tag{3.1}$$

并把它称之为均方误差最小准则. 用这种准则所确定的预报估值

$\hat{x}_{k+m}$ 满足

$$\mathscr{B}^*(k, m) = \min \mathscr{B}(k, m)$$
$$= \min_{F \in \mathscr{F}} E\{|x_{k+m} - F(x_k, x_{k-1}, \cdots, x_{k-n})|^2\}$$

其中 $\mathscr{F}$ 是某种函数的族，如线性函数的族、连续函数的族等等.

我们把用上述准则所确定的预报估值 $\hat{x}_{k+m}$ 称为最小均方误差预报. 如果估值 $\hat{x}_{k+m}$ 还具有无偏性，即如果有

$$E\{\hat{x}_{k+m}\} = x_{k+m}$$

则称 $\hat{x}_{k+m}$ 为 x_{k+m} 的最小方差预报.

定理 3.1 x_{k+m} 依据资料 $\{x_k, x_{k-1}, \cdots, x_{k-m}\}$ 的最小均方误差预报 $\hat{x}_{k+m}$ 必满足

$$\hat{x}_{k+m} = E\{x_{k+m}|x_k, x_{k-1}, \cdots, x_{k-n}\} \quad \text{a.s}$$

$E\{\cdot|x_k, x_{k-1}, \cdots, x_{k-n}\}$ 表示在条件 $\{x_k, x_{k-1}, \cdots, x_{k-n}\}$ 下的条件数学期望(均值).

证明 为使符号简便计，令 $\mathbf{X}_n^k = \{x_k, x_{k-1}, \cdots, x_{k-n}\}$. 由条件数学期望的性质，我们有

$$E\{|x_{k+m} - F(\mathbf{X}_n^k)|^2\} = E\{E\{|x_{k+m} - F(\mathbf{X}_n^k)|^2|\mathbf{X}_n^k\}\}$$

而

$$E\{|x_{k+m} - F(\mathbf{X}_n^k)|^2|\mathbf{X}_n^k\}$$
$$= E\{|x_{k+m} - E\{x_{k+m}|\mathbf{X}_n^k\} + E\{x_{k+m}|\mathbf{X}_n^k\} - F(\mathbf{X}_n^k)|^2|\mathbf{X}_n^k\}$$
$$= E\{|x_{k+m} - E\{x_{k+m}|\mathbf{X}_n^k\}|^2|\mathbf{X}_n^k\}$$
$$-2E\{(x_{k+m} - E\{x_{k+m}|\mathbf{X}_n^k\})(E\{x_{k+m}|\mathbf{X}_n^k\} - F(\mathbf{X}_n^k))|\mathbf{X}_n^k\}$$
$$+ E\{|E\{x_{k+m}|\mathbf{X}_n^k\} - F(x_n^k)|^2|\mathbf{X}_n^k\}$$

但是在 $F(\mathbf{X}_n^k)$ 是可测函数的假定下，有

$$E\{(x_{k+m} - E\{x_{k+m}|\mathbf{X}_n^k\})(E\{x_{k+m}|\mathbf{X}_n^k\} - F(\mathbf{X}_n^k))|\mathbf{X}_n^k\}$$
$$= (E\{x_{k+m}|\mathbf{X}_n^k\} - F\{\mathbf{X}_n^k\})E\{(x_{k+m} - E\{x_{k+m}|\mathbf{X}_n^k\})|\mathbf{X}_n^k\}$$
$$= (E\{x_{k+m}|\mathbf{X}_n^k\} - F(\mathbf{X}_n^k))\{E\{x_{k+m}|\mathbf{X}_n^k\} - E\{x_{k+m}|\mathbf{X}_n^k\}\}$$
$$= 0$$

所以

$$E\{|x_{k+m}-F(\mathbf{X}_n^k)|^2\}=E\{|x_{k+m}-E\{x_{k+m}|\mathbf{X}_n^k\}|^2\}$$
$$+E\{|E\{x_{k+m}|\mathbf{X}_n^k\}-F(\mathbf{X}_n^k)|^2\}$$

由此可见，只有当

$$F(\mathbf{X}_n^k)=E\{x_{k+m}|\mathbf{X}_n^k\}\quad \text{a.s}$$

时才能使

$$E\{|x_{k+m}-F(\mathbf{X}_n^k)|^2\}=\min \qquad \blacksquare$$

这个定理说明了，在可测函数类中，x_{k+m} 依据资料 $\mathbf{X}_n^k$ 的最小均方误差预报值是条件均值 $E\{x_{k+m}|\mathbf{X}_n^k\}$. 并且显然有

$$E\{\hat{x}_{k+m}\}=E\{E\{x_{k+m}|\mathbf{X}_n^k\}\}=x_{k+m}$$

遗憾的是，在一般情形，条件均值

$$E\{x_{k+m}|\mathbf{X}_n^k\}$$

的具体表达式是很难找到的. 所以为寻求预报问题的有实用价值的解法，在大多数情形，还必须另寻途径.

2. 线性预报

如果时间序列$\{x_k\}$服从具有零均值的正态分布，则预报问题的求解就变得简单了. 因为当 $(x_{k+m}, x_k, x_{k-1}, \cdots, x_{k-n})$ 的联合分布为正态分布时，我们有

$$E\{x_{k+m}|x_k, x_{k-1}, \cdots, x_{k-n}\}=a_0x_k+a_1x_{k-1}+\cdots$$
$$+a_nx_{k-n}$$

其中 $a_0, a_1, \cdots, a_n$ 是与m和n有关的常数. 在这种情形，为确定 $\hat{x}_{k+m}$，只需得出 $a_0, a_1, \cdots, a_n$ 的估值就行了.

在一般的情形，虽然 $\{x_k\}$ 不具有正态分布，但我们仍可以考虑满足条件

$$E\{|x_{k+m}-a_0x_k-a_1x_{k-1}-\cdots-a_nx_{k-n}|^2\}=\min$$

的关于 x_{k+m} 的预报估值

$$\hat{x}_{k+m}=a_0x_k+a_1x_{k-1}+\cdots+a_nx_{k-n} \tag{3.2}$$

我们称这样的估值为线性最小均方误差预报估值. (3.2) 式称为线性最小均方误差预报公式.

定理 3.2 如果$\{x_k\}$是一个零均值的平稳随机序列. 并且其

协方差函数的值 $r(0), r(1), \cdots, r(m), r(m+1), \cdots, r(m+n)$ 是已知的,则其线性预报公式中的未知参数 $a_0, a_1, \cdots, a_n$ 可由下述方程组确定

$$\Gamma \mathbf{a} = \mathbf{b}$$

其中

$$\Gamma = \begin{pmatrix} r(0) & r(1) \cdots r(n) \\ r(1) & r(0) \cdots r(n-1) \\ \vdots & \vdots \quad\quad \vdots \\ r(n) & r(n-1) \cdots r(0) \end{pmatrix}$$

$$\mathbf{a} = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} r(m) \\ r(m+1) \\ \vdots \\ r(m+n) \end{pmatrix}$$

证明 置

$$\mathscr{B}(a_0, a_1, \cdots, a_n) = E\left\{\left|x_{k+m} - \sum_{i=0}^{n} a_i x_{k-i}\right|^2\right\}$$

我们的目的是寻求 $a_0, a_1, \cdots, a_n$, 使上式达到最小. 为此,应用初等微分学的方法,求 $\mathscr{B}(a_0, \cdots, a_n)$ 关于 $a_0, a_1, \cdots, a_n$ 的偏导数并令它们为 0, 即得

$$\frac{\partial \mathscr{B}(a_0, a_1, \cdots, a_n)}{\partial a_j} = 2E\left\{\left(x_{k+m} - \sum_{i=0}^{n} a_i x_{k-i}\right)(-x_{k-j})\right\}$$
$$= 0 \qquad j = 0, 1, \cdots, n$$

或者是

$$\begin{aligned} r(0)a_0 + r(1)a_1 + \cdots + r(n)a_n &= r(m) \\ r(1)a_0 + r(0)a_1 + \cdots + r(n-1)a_n &= r(m+1) \\ \vdots \qquad \vdots \qquad\qquad \vdots \qquad & \qquad \vdots \\ r(n)a_0 + r(n-1)a_1 + \cdots + r(0)a_n &= r(m+n) \end{aligned}$$

写成矩阵向量形式,即有

$$\Gamma \mathbf{a} = \mathbf{b}$$ 

上述定理说明了,在序列为平稳的情形,只要相应的协方差函数的值为已知,那么依赖"有限"过去的资料 $\{x_k, x_{k-1}, \cdots, x_{k-n}\}$ (n 有限)的预报问题是容易解决的,预报公式中未知参数的估值为

$$\hat{\mathbf{a}} = \Gamma^{-1}\mathbf{b}$$

§ 3.2 瓦尔德分解

1. 具有二阶矩的随机变量所形成的希耳伯特(Hilbert)空间

设H是所有具有二阶矩的并且均值为零的随机变量形成的集合. 对于任何的 $\mathbf{x}, \mathbf{y} \in H$, 令

$$\langle \mathbf{x}, \mathbf{y} \rangle = E\{\mathbf{x} \cdot \bar{\mathbf{y}}\} \tag{3.3}$$

其中 $\bar{\mathbf{y}}$ 表示 $\mathbf{y}$ 的共轭复数. 不难验证,(3.3)式定义了H上的一个内积. 在这个内积之下,H形成了一个希耳伯特空间, 其度量(距离)是 $\|\mathbf{x}-\mathbf{y}\| = E\{|\mathbf{x}-\mathbf{y}|^2\}^{\frac{1}{2}}$. 在这个空间中,两个元素 $\mathbf{x}$ 和 $\mathbf{y}$ 是直交的,如果

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0$$

下述的定理说明了时间序列的线性最小均方误差预报的几何意义.

定理 3.3 设 $\{x_k\}$ 是一个零均值的具有二阶矩的随机序列, 则 $\hat{F}(\mathbf{X}_n^k)$ 是 x_{k+m} 的最小均方误差线性预报的充分与必要条件是对任何的 $y \in L\{\mathbf{X}_n^k\}$ 皆有

$$\langle x_{k+m} - \hat{F}(\mathbf{X}_n^k), y \rangle = 0 \tag{3.4}$$

其中$\mathbf{X}_n^k = \{x_k, x_{k-1}, \cdots, x_{k-n}\}$. $L(\mathbf{X}_n^k)$是由形如$\left\{\sum_{k=0}^{n} \alpha_i x_{k-i} \,\middle|\, \alpha_0, \alpha_1, \cdots, \alpha_n \text{ 是任意常数}\right\}$ 的集合所形成的子空间.

证明 为简便计,我们只考虑实数的情形. 由于 $\hat{F}(\mathbf{X}_n^k)$ 是 $\mathbf{X}_n^k = \{x_n, x_{k-1}, \cdots, x_{k-n}\}$ 的线性函数,故有 $F(\mathbf{X}_n^k) \in L(\mathbf{X}_n^k)$.

充分性：设对一切 $y\in L(\mathbf{X}_n^k)$ 皆有

$$\langle x_{k+m}-\hat{F}(\mathbf{X}_n^k),\ y\rangle=0$$

任取 $y\in L(\mathbf{X}_n^k)$，令 $x=y-\hat{F}(\mathbf{X}_n^k)$，即

$$y=\hat{F}(\mathbf{X}_n^k)+x$$

则有

$$\langle x_{k+m}-\hat{F}(\mathbf{X}_n^k),\ x\rangle=0$$

所以

$$\begin{aligned}\|x_{k+m}-y\|^2&=\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)-x\|^2\\&=\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)\|^2+\|x\|^2\geqslant\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)\|^2\end{aligned}$$

从而

$$\min_{y\in L}(\mathbf{X}_n^k)\|x_{k+m}-y\|^2=\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)\|^2$$

必要性：用反证法，设有 $\omega\in L(\mathbf{X}_n^k)$，使

$$\langle x_{k+m}-\hat{F}(\mathbf{X}_n^k),\ \omega\rangle=\alpha\|\omega\|^2\neq 0$$

此处 α 是不为零的常数，置

$$y=\hat{F}(\mathbf{X}_n^k)+\alpha\omega$$

显然，$y\in L(\mathbf{X}_n^k)$，而且

$$\begin{aligned}\|x_{k+m}-y\|^2&=\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)-\alpha\omega\|^2\\&=\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)\|^2+\alpha^2\|\omega\|^2-2\langle x_{k+m}-\hat{F}(\mathbf{X}_n^k),\alpha\omega\rangle\\&=\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)\|^2-\alpha^2\|\omega\|^2\end{aligned}$$

即有 $y\in L(\mathbf{X}_n^k)$，使

$$\|x_{k+m}-y\|<\|x_{k+m}-\hat{F}(\mathbf{X}_n^k)\|$$

所以 $\hat{F}(\mathbf{X}_n^k)$ 不是最小均方误差线性估计. ■

根据这个定理，我们可以把 x_{k+m} 的最小均方误差预报 $\hat{x}_{k+m}$，称为 x_{k+m} 在 $L(\mathbf{X}_n^k)$ 上的投影. 不难发现，当 n 无限时，上述定理仍成立.

2. 瓦尔德（Wold）分解

设 $\{x_k\}$ 是一个零均值的平稳时间序列，置

$$H_k=L\{x_k,\ x_{k-1},\cdots\}$$

即 H_k 是由元素集 $\{x_k,\ x_{k-1},\ \cdots\}$ 所张成的线性子空间. $\hat{x}_{k+m}$

是依据资料 $\{x_k, x_{k-1}, \cdots\}$ 对 x_{k+m} 的最小均方误差线性预报．令

$$\sigma_m^2 = E\{|x_{k+m} - \hat{x}_{k+m}|^2\} \tag{3.5}$$

则显然有

1) $H_k \subset H_l$，当 $k \leqslant l$；

2) $\sigma_1^2 \leqslant \sigma_2^2 \leqslant \cdots \leqslant \sigma_m^2 \leqslant \cdots$．

如果 $\sigma_1^2 > 0$，则平稳序列 $\{x_k\}$ 称为是正则的（非确定性的），如果 $\sigma_1^2 = 0$，则 $\{x_k\}$ 称为是奇异的(确定性的)．

因为有

$$\sigma_1^2 \leqslant \sigma_m^2 \qquad m \geqslant 1$$

所以如果 $\sigma_1^2 > 0$，则对于一切 $m \geqslant 1$ 皆有 $\sigma_m^2 > 0$．另一方面，如果 $\sigma_1^2 = 0$，则有

$$x_{k+1} = \hat{x}_{k+1} \in H_k$$

由此得出 $H_{k+1} \subset H_k$，从而 $H_{k+1} = H_k$．这就是说对于一切 $k \geqslant 1$，H_k 皆相同，从而必有

$$0 = \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2 = \cdots$$

总之我们有：

定理 3.4 数列 $\{\sigma_m^2\}$ 中的每项或者全为正(此时 $\{x_k\}$ 是正则的)，或者全为零(此时 $\{x_k\}$ 是奇异的)．

瓦尔德在 1938 年证明了，任何一个平稳序列都能够分解成为相互无关的两部分，一部分是正则的，另一部分是奇异的．确切地说，我们有：

定理 3.5(瓦尔德分解定理) 任何平稳序列 $\{x_k\}$ 皆能够分解成为

$$x_k = u_k + v_k \tag{3.6}$$

其中 $\{u_k\}$ 和 $\{v_k\}$ 具有如下性质：

1) $\{u_k\}$ 和 $\{v_k\}$ 是相互无关的．

2) u_k 是正则的，并且具有单边滑动平均的形式

$$u_k = \sum_{i=0}^{\infty} c_i e_{k-i} \tag{3.7}$$

其中 $c_0=1$，$\sum_{i=0}^{\infty} c_i^2<+\infty$，$\{e_k\}$ 是零均值的白噪声。而且 $\{e_k\}$ 与 $\{v_k\}$ 是无关的.

3) v_k 是奇异的.

证明 仅考虑实数的情形. 设 $\hat{x}_k$ 表示 x_k 基于资料 x_{k-1}, x_{k-2}, $\cdots$的最小均方误差预报，令

$$e_k=x_k-\hat{x}_k$$

显然，e_k 与 H_{k-1}中的每个元素皆正交，这一事实被称之为 e_k 与 H_{k-1} 正交，记作 $e_k\perp H_{k-1}$. 对于任何的正整数 k, l: $k<l$, 则由于 $H_k\subset H_{l-1}$ 及 $e_l\in H_l$，故有 $e_l\perp H_k$，从而 $\langle e_k, e_l\rangle=0$，并且 $E\{e_k\}=E\{x_k-\hat{x}_k\}=0$，所以，$\{e_k\}$ 是零均值的白噪声.

用 $H_k(e)$ 表示由元素 $e_k, e_{k-1},\cdots$ 所张成的子空间. 元素 x_k 在 $H_k(e)$ 上的投影记成 u_k. 于是 $u_k\in H_k(e)$，并且对一切 $y\in H_k(e)$

$$\langle x_k-u_k, y\rangle=0$$

显然 u_k 可有形式

$$u_k=\sum_{i=0}^{\infty} c_i e_{k-i} \tag{3.8}$$

令 $\sigma_k^2(e)=E\{e_k^2\}=\|e_k\|^2$. 注意到

$$\langle x_k, e_l\rangle=\langle u_k, e_l\rangle \qquad 对一切\ l\geqslant 1$$

所以我们有

$$c_i=\frac{1}{\sigma_{k-i}^2(e)}E\{x_k e_{k-i}\}$$

特别地

$$c_0=\frac{1}{\sigma_k^2(e)}E\{x_k e_k\}=\frac{1}{\sigma_k^2(e)}E\{(\hat{x}_k+e_k)e_k\}$$

$$=\frac{1}{\sigma_k^2(e)}E\{e_k^2\}=1$$

不妨设 $\sigma_k^2(e)>0$，如不然 $\sigma_k^2(e)=0$，则序列 $\{x_k\}$ 自身是奇异的，只要取 $u_k=0$，$v_k=x_k$，定理的结论自然成立.

总之，我们说明了 $\{u_k\}$ 是正则的并有单边滑动和的形式

$$u_k = \sum_{i=0}^{\infty} c_i e_{k-i} \qquad c_0 = 1$$

置

$$v_k = x_k - u_k \tag{3.9}$$

则对于 $l \leqslant k$，我们有

$$\langle v_k, e_l \rangle = \langle x_k - u_k, e_l \rangle = 0$$

对于 $l > k$，由于 $v_k \in H_k$，$e_l \perp H_k$，所以 $e_l \perp v_k$. 总之，对一切 l, k 皆有 $e_l \perp v_k$. 从而对一切 l，k 皆有 $u_l \perp v_k$. 即 $\{u_k\}$ 与 $\{v_k\}$ 无关. 并且

$$\|x_k\|^2 = \|u_k\|^2 + \|v_k\|^2$$

注意到 $\{e_k\}$ 的平稳性，故得 $\sigma_k^2(e) = \sigma_e^2$ 对一切 k. 所以有

$$\sum_{i=0}^{\infty} c_i^2 = \frac{1}{\sigma_e^2}\|u_k^2\| \leqslant \frac{1}{\sigma_e^2}\|x_k\|^2 < +\infty$$

最后来证明 $\{v_k\}$ 是奇异的. 为此注意到

$$v_k \in H_k = H_{k-1} \oplus \{e_k\}$$

此处 $\{e_k\}$ 表示由 e_k 所生成的子空间. 但已知 $v_k \perp e_k$，所以必有 $v_k \in H_{k-1}$. 重复上述讨论又可得 $v_k \in H_{k-2}$，一般地可得出：对 $l \geqslant 0$，$v_k \in H_{k-l}$. 但由等式(3.8)和(3.9)可见

$$H_k = H_k(v) \oplus H_k(e)$$

此处 $H_k(v)$ 是由 $\{v_k, v_{k-1}, \cdots\}$ 所张成的子空间，$H_k(e)$ 是由 $\{e_k, e_{k-1}, \cdots\}$ 所张成的子空间. 但已知 $v_k \in H_{k-l}$，对一切 $l \geqslant 0$. 而 $v_k \bar{\in} H_k(e)$，所以必有 $v_k \in H_{k-1}(v)$. 即

$$H_k(v) = H_{k-1}(v)$$

由此即可得出 v_k 的奇异性. ■

进一步，还可以证明下述的定理：

定理 3.6 设 $f(\lambda)$ 是平稳时间序列 $\{x_k\}$ 的谱密度函数，如果它满足

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda > -\infty \tag{3.10}$$

则必存在数列 $c_0, c_1, \cdots, c_n, \cdots, c_0$ 是正实数，

$$\sum_{i=0}^{\infty} |c_i|^2 < +\infty$$

使函数

$$G(z) = \sum_{k=0}^{\infty} c_k z^k$$

在单位圆内没有零点，而且

$$f(\lambda) = |G(e^{-2\pi i\lambda})|^2$$

进一步可以看出，如果 $\{x_k\}$ 的谱密度 $f(\lambda)$ 满足条件(3.10)，则 x_k 必可表示成

$$x_k = \sum_{i=0}^{\infty} c_i e_{k-i}$$

其中 $\{e_k\}$ 是零均值的白噪声，$E\{e_k^2\} = 1$.

§3.3 平稳时间序列预报的维纳柯尔莫哥洛夫方法

1. 维纳方法

对于平稳时间序列 $\{x_k\}$ 的线性预报问题，用希耳伯特空间的语言，就是要寻求 x_{k+m} 在空间 H_k 上的投影，此处 H_k 是由$\{x_k, x_{k-1}, \cdots\}$ 所张成的子空间. 由此不难看出，$\hat{x}_{k+m}$ 应该具有形式

$$\hat{x}_{k+m} = \sum_{i=0}^{\infty} a_i x_{k-i} \tag{3.11}$$

其中系数 $a_0, a_1, \cdots$ 应使下述预报准则取最小

$$\mathscr{B}(k, m) = E\{|x_{k+m} - \hat{x}_{k+m}|^2\}$$

$$= r(0) - 2\sum_{i=0}^{\infty} a_i r(m+i) + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_i a_j r(i-j)$$

此处，$r(k)$ 是 $\{x_k\}$ 的协方差函数. 应用微分方法可以看出使

$\mathscr{B}(k, m)$ 取最小值的 $a_0, a_1, \cdots$ 应该满足

$$\sum_{j=0}^{\infty} a_j r(i-j) = r(m+i) \quad i = 0, 1, 2, \cdots \tag{3.12}$$

这组方程称之为维纳-霍夫（Wiener-Hopf）方程．当资料个数为有限时，即已知 $\{x_k, x_{k-1}, \cdots, x_{k-n}\} = \mathbf{X}_n^k$ 时，上述方程变成了

$$\sum_{j=0}^{n} a_j r(i-j) = r(m+i) \quad i = 0, 1, 2, \cdots, n$$

当 $\{x_k\}$ 的协方差函数已知时，通过上述方程就可完全确定 a_0, $a_1, \cdots$．然而在 n 为无穷的情形，这种途径就行不通了．

为克服上述困难，按照维纳的途径，把预报问题的求解转化到频域中来考虑．为此，回忆到平稳序列的频谱展开式

$$x_{k+m} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i(k+m)\lambda} dZ(\lambda)$$

其中直交增量过程 $\left\{Z(\lambda), -\frac{1}{2} \leqslant \lambda \leqslant \frac{1}{2}\right\}$ 满足 $E\{|dZ(\lambda)|^2\} = f(\lambda)d\lambda$，$f(\lambda)$ 是谱密度函数．

进一步，只要置

$$A(e^{-2\pi i\lambda}) = \sum_{j=0}^{\infty} a_j e^{-2\pi ij\lambda}$$

就可以把(3.11)式写成下述形式

$$\hat{x}_{k+m} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda} A(e^{-2\pi i\lambda}) dZ(\lambda)$$

于是有

$$x_{k+m} - \hat{x}_{k+m} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}\{e^{2\pi im\lambda} - A(e^{-2\pi i\lambda})\} dZ(\lambda)$$

从而

$$\begin{aligned} \mathscr{B}(k, m) &= E\{|x_{k+m} - \hat{x}_{k+m}|^2\} \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |e^{2\pi im\lambda} - A(e^{-2\pi i\lambda})|^2 f(\lambda) d\lambda \end{aligned}$$

在讨论平稳序列的预报问题时，根据瓦尔德分解，我们只需考虑正则的序列就够了．根据定理 3.6，它的一个充分条件是

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda)d\lambda > -\infty \tag{3.13}$$

其中 $f(\lambda)$ 是序列的谱密度．此时

$$f(\lambda) = |G(e^{-2\pi i\lambda})|^2$$

注意到 $f(\lambda) \geq 0$，故在正则化条件 $G(0) = 1$ 的情况下，可以把 $f(\lambda)$ 写成

$$f(\lambda) = c|G(e^{-2\pi i\lambda})|^2 \tag{3.14}$$

其中 $c > 0$。以下我们恒假定时间序列 $\{x_k\}$ 满足 (3.13) 式，于是

$$\mathscr{B}(k, m) = c\int_{-\frac{1}{2}}^{\frac{1}{2}} |e^{2\pi im\lambda}G(e^{-2\pi i\lambda}) - A(e^{-2\pi i\lambda})G(e^{-2\pi i\lambda})|^2 d\lambda$$

为方便计，引入下列记号

设 $F(z)$ 是一个解析函数，对它进行罗郎 (Laurent) 展开

$$F(z) = \sum_{k=-\infty}^{+\infty} f_k z^k$$

令

$$[F(z)]_+ = \sum_{k=0}^{\infty} f_k z^k \quad [F(z)]_- = \sum_{k=-\infty}^{-1} f_k z^k$$

即 $[F(z)]_+$ 是 $F(z)$ 的罗郎级数的非负指数部分，$[F(z)]_-$ 是负指数部分．对于函数 $\bar{z}^m G(z)$ 应用上述记号，我们有

$$\bar{z}^m G(z) = G_1(z) + G_2(z)$$

其中

$$G_1(z) = [\bar{z}^m G(z)]_+ \quad G_2(z) = [\bar{z}^m G(z)]_-$$

于是

$$e^{2\pi im\lambda}G(e^{-2\pi i\lambda}) = G_1(e^{-2\pi i\lambda}) + G_2(e^{-2\pi i\lambda})$$

其中

$$G_1(e^{-2\pi i\lambda}) = \sum_{k=0}^{\infty} g_k e^{-2\pi ik\lambda}$$

$$G_2(e^{-2\pi i\lambda}) = \sum_{k=-\infty}^{-1} g_k e^{-2\pi ik\lambda}$$

再注意到 $A(e^{-2\pi i\lambda})$ 和 $G(e^{-2\pi i\lambda})$ 的形式

$$A(e^{-2\pi i\lambda})=\sum_{k=0}^{\infty}a_k e^{-2\pi ik\lambda}$$

$$G(e^{-2\pi i\lambda})=\sum_{k=0}^{\infty}c_k e^{-2\pi ik\lambda}$$

所以必有

$$\int_{-\frac{1}{2}}^{\frac{1}{2}}[G_1(e^{-2\pi i\lambda})-A(e^{-2\pi i\lambda})G(e^{-2\pi i\lambda})]\overline{G_2(e^{-2\pi i\lambda})}d\lambda=0$$

从而

$$\mathscr{B}(k,m)=c\int_{-\frac{1}{2}}^{\frac{1}{2}}|G_2(e^{-2\pi i\lambda})|^2d\lambda+c\int_{-\frac{1}{2}}^{\frac{1}{2}}|G_1(e^{-2\pi i\lambda})-A(e^{-2\pi i\lambda})G(e^{-2\pi i\lambda})|^2d\lambda$$

由上式可见,右端的两项皆是非负的,而且第一项与 $A(e^{-2\pi i\lambda})$ 无关. 为选取 $A(z)$ 使 $\mathscr{B}(k,m)$ 最小,只需取

$$A(z)=\frac{G_1(z)}{G(z)}=\frac{[z^{-m}G(z)]_+}{G(z)}$$

由此可见,只要能够求出 $G(z)$,那么预报问题就解决了. 此时 $\hat{x}_{k+m}$ 的频谱展开式是

$$\hat{x}_{k+m}=\int_{-\frac{1}{2}}^{\frac{1}{2}}e^{2\pi ik\lambda}A(e^{-2\pi i\lambda})dZ(\lambda)$$

而它在时域中的表达式是

$$\hat{x}_{k+m}=A(q^{-1})x_k$$

预报的均方误差是

$$\sigma_m^2=E\{|x_{k+m}-\hat{x}_{k+m}|^2\}=c\int_{-\frac{1}{2}}^{\frac{1}{2}}|G_2(e^{-2\pi i\lambda})|^2d\lambda$$

2. 柯尔莫哥洛夫 (Kolmogorov) 方法

柯尔莫哥洛夫方法也是从表达式

$$\hat{x}_{k+m}=\sum_{i=0}^{\infty}a_i x_{k-i}\tag{3.15}$$

出发．由于我们假定了序列 $\{x_k\}$ 的谱密度 $f(\lambda)$ 满足条件

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda > -\infty$$

故 x_k 可以表示成单边滑动和的形式

$$x_k = \sum_{j=0}^{\infty} c_j e_{k-j} \tag{3.16}$$

其中 $\{e_k\}$ 是零均值的白噪声序列，$E\{e_k^2\} = \sigma_e^2$ 把(3.16)式代入(3.15)式后，经整理可得

$$\hat{x}_{k+m} = \sum_{j=0}^{\infty} b_j e_{k-j} \tag{3.17}$$

由于 $\mathrm{var}\{\hat{x}_{k+m}\} \leqslant \mathrm{var}\{x_{k+m}\} < +\infty$，所以必有

$$\sum_{j=0}^{\infty} b_j^2 < +\infty$$

另一方面应用(3.16)式可得

$$\begin{aligned} x_{k+m} &= \sum_{j=0}^{\infty} c_j e_{k+m-j} \\ &= \sum_{j=0}^{m-1} c_j e_{k+m-j} + \sum_{j=m}^{\infty} c_j e_{k+m-j} \end{aligned} \tag{3.18}$$

其中 $\sum_{j=0}^{m-1} c_j e_{k+m-j}$ 是 $e_{k+m}, e_{k+m-1}, \cdots, e_{k+1}$ 的线性组合．这里所涉及的是 e_k 的未来值，所以是不可预报的．而 $\sum_{j=0}^{m-1} c_j e_{k+m-j}$ 与 $\sum_{j=m}^{\infty} c_j e_{k+m-j}$ 是不相关的．故必有

$$\hat{x}_{k+m} = \sum_{j=m}^{\infty} c_j e_{k+m-j} = \sum_{j=0}^{\infty} c_{j+m} e_{k-j} \tag{3.19}$$

而

$$x_{k+m} - \hat{x}_{k+m} = \sum_{j=0}^{m-1} c_j e_{k+m-j}$$

我们可以直接说明(3.17)式中的 b_i 取为 c_{i+m} 时,必可使

$$\mathscr{B}(k, m) = E\{|x_{k+m} - \hat{x}_{k+m}|^2\} = \min$$

事实上,把(3.18)和(3.17)式代入 $\mathscr{B}(k, m)$ 中得

$$\mathscr{B}(k, m) = E\left\{\left[\sum_{j=0}^{m-1} c_j e_{k+m-j} + \sum_{j=0}^{\infty} (c_{j+m} - b_j) e_{k-j}\right]^2\right\}$$

$$= \sigma_e^2 \left\{\left(\sum_{j=0}^{m-1} c_j^2\right) + \sum_{i=0}^{\infty} (c_{j+m} - b_j)^2\right\}$$

由此可见,只要取 $b_i = c_{i+m}$, 才能使 $\mathscr{B}(k, m)$最小,并且此时有

$$\sigma_m^2 = \min \mathscr{B}(k, m) = \sigma_e^2 \left(\sum_{j=0}^{m-1} c_j^2\right)$$

由上述讨论可以看出,只要能够确定 $c_i(i = 0, 1, \cdots)$,预报问题就解决了. 但我们知道,这些 $c_i(j = 0, 1, \cdots)$ 满足

$$G(z) = \sum_{j=0}^{\infty} c_j z^j$$

而

$$f(\lambda) = |G(e^{-2\pi i \lambda})|^2$$

所以关键在于求出这样的函数 $G(z)$.

当然,最终我们可以、而且也应该用 $x_k, x_{k-1}, \cdots$ 来表示 $\hat{x}_{k+m}$. 事实上,如果置

$$G^{(m)}(z) = \sum_{j=0}^{\infty} c_{j+m} z^j$$

则

$$\hat{x}_{k+m} = G^{(m)}(q^{-1}) e_k$$

此外由(3.16)式有

$$x_k = G(q^{-1}) e_k$$

即

$$e_k = G(q^{-1})^{-1} x_k$$

所以

$$\hat{x}_{k+m} = G^{(m)}(q^{-1}) G(q^{-1})^{-1} x_k = A(q^{-1}) x_k$$

其中 $A(q^{-1}) = G^{(m)}(q^{-1})G(q^{-1})^{-1}$. 利用前面所介绍的记号，我们有

$$G^{(m)}(z) = [z^{-m}G(z)]_+$$

所以有

$$A(z) = \frac{[z^{-m}G(z)]_+}{G(z)}$$

这说明了柯尔莫哥洛夫方法和维纳方法解出的结果是一样的.

3. 谱密度函数的因子分解

从前面的讨论可以看出，用维纳和柯尔莫哥洛夫方法解决预报问题，最终归结为寻求函数 $G(z)$，使

$$f(\lambda) = |G(e^{-2\pi i\lambda})|^2 \tag{3.20}$$

注意到谱密度函数可以用协方差函数 $r(k)$ 表示出来，即

$$f(\lambda) = \sum_{k=-\infty}^{+\infty} r(k)e^{-2\pi ik\lambda}$$

因为 $r(k) = r(-k)$，所以 $f(-\lambda) = f(\lambda)$，故 $\log f(\lambda)$ 也是一个偶函数. 令 $z = e^{-2\pi i\lambda}$，$f^*(z) = f(\lambda)$对 $\log f^*(\lambda)$ 进行罗郎展开

$$\log f^*(z) = \sum_{j=-\infty}^{+\infty} g_j z^j$$

于是我们有

$$\begin{aligned} f^*(z) &= \exp\left\{\sum_{j=-\infty}^{+\infty} g_j z^j\right\} \\ &= e^{g_0}\exp\left\{\sum_{j=1}^{\infty} g_j z^j\right\}\exp\left\{\sum_{j=-\infty}^{-1} g_j z^j\right\} \end{aligned}$$

另一方面

$$\log f(\lambda) = \sum_{j=-\infty}^{+\infty} g_j e^{-2\pi ij\lambda}$$

所以 $g_j(j = 0, \pm 1, \pm 2, \cdots)$ 是偶函数 $\log f(\lambda)$ 的傅里叶系数，从而有 $g_{-j} = g_j$. 由此可见，如果令

$$G(z) = \exp\left\{\sum_{j=1}^{\infty} g_j z^j\right\}$$

则我们有

$$f(\lambda) = e^{g_0}|G(e^{-2\pi i\lambda})|^2$$

这就是说,我们已把谱密度函数表示成了(3.20)式的形式，其中常数因子 e^{g_0} 存在,并不会有什么影响.

§ 3.4 时间序列预报的博克斯-詹金斯方法

博克斯-詹金斯 Box-Jenkins 方法所得出的仍然是最小均方误差预报. 它所能够处理的是具有有限多个参量的 ARMA 类型的序列. 对于平稳序列的情形，ARMA 序列与具有有理谱密度的序列之间存在着一一对应的关系. 而具有有理谱密度的过程是具有一定的普遍性的. 博克斯-詹金斯方法的最大优越性在于,它能够实现递推预报,即当预报的步长从 m 改变成 $m+1$ 时,预报算法不必从头重复做起. 这将给算法的实际应用带来方便.

设 $\{X_k\}$ 满足 ARMA(n, p) 模型

$$A(q^{-1})x_k = C(q^{-1})e_k \tag{3.21}$$

其中 $\{e_k\}$ 是具有零均值和公共方差 σ_e^2 的独立随机变量序列.

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$

$$C(q^{-1}) = 1 + cq^{-1} + \cdots + c_pq^{-p}$$

多项式 $A(z), C(z)$ 无公因子,并且它们的零点皆在单位圆外.为方便计,我们把 $\hat{x}_{k+m}$ 改写成为 $\hat{x}_k(m)$, 即

$$\hat{x}_k(m) = E\{x_{k+m}|x_k, x_{k-1}, \cdots\}$$

考虑模型

$$A(q^{-1})x_{k+1} = C(q^{-1})e_{k+1}$$

即

$$x_{k+1} + a_1x_k + \cdots + a_nx_{k-n+1} = e_{k+1} + ce_k + \cdots + c_pe_{k-p+1} \tag{3.22}$$

令 $\mathbf{X}_k = \{x_k, x_{k-1}, \cdots\}$. 对上式两端取条件数学期望 $E\{\cdot|\mathbf{X}_k\}$,

我们有

$$\hat{x}_k(1) + a_1x_k + \cdots + a_nx_{k-n+1} = E\{e_{k+1}|\mathbf{X_k}\} + c_1E\{e_k|\mathbf{X_k}\}^+ + \cdots + c_pE\{e_{k-p+1}|\mathbf{X_k}\} \quad (3.23)$$

另一方面注意到,对一切 k

$$x_k = A(q^{-1})^{-1}C(q^{-1})e_k$$

即 x_k 可由 $e_k, e_{k-1}, \cdots$线性表示。所以

$$E\{e_{k+1}|\mathbf{X_k}\} = E\{e_{k+1}|e_k, e_{k-1}, \cdots\} = 0$$

$$E\{e_{k-h}|\mathbf{X_k}\} = e_{k-h} \qquad h \geqslant 0$$

从而(3.23)式可以写成

$$\hat{x}_k(1) + a_1x_k + \cdots + a_nx_{k-n+1} = c_1e_k + \cdots + c_pe_{k-p+1} \quad (3.24)$$

由(3.22)式中减去(3.24)式得

$$x_{k+1} - \hat{x}_k(1) = e_{k+1}$$

完全类似地有

$$e_k = x_k - \hat{x}_{k-1}(1) \qquad e_{k-1} = x_{k-1} - \hat{x}_{k-2}(1) \ \cdots$$

所以

$$\hat{x}_k(1) + a_1x_k + \cdots + a_nx_{k-n+1} = c(x_k - \hat{x}_{k-1}(1)) + \cdots + \cdots + c_p(x_{k-p+1} - \hat{x}_{k-p}(1))$$

即当 $n \geqslant p$ 时,有

$$\hat{x}_k(1) + c\hat{x}_{k-1}(1) + \cdots + c_p\hat{x}_{x-p}(1) = (c - a_1)x_k + \cdots + (c_p - a_p)x_{k-p+1} - a_{p+1}x_{k-p} - a_nx_{k-n+1} \quad (3.25)$$

完全类似地可以得出

$$\begin{aligned}\hat{x}_k(2) + a_1\hat{x}_k(1) + a_2x_k + \cdots + a_mx_{k-n+2} \\ = c_2e_k + \cdots + c_pe_{k-p+2} \\ = c_2[x_k - \hat{x}_{k-1}(1)] + \cdots + c_p[x_{k-p+2} - \hat{x}_{k-p+1}(1)]\end{aligned}$$

或

$$\hat{x}_k(2) + a_1\hat{x}_k(1) + c_2\hat{x}_{k-1}(1) + \cdots + c_p\hat{x}_{k-p+1}(1) = (c_2 - a_2)x_k + \cdots + (c_p - a_p)x_{k-p+2} - a_{p+1}x_{k-p+1} - \cdots - a_nx_{k-n+2} \quad (3.26)$$

再重复上述的手续，也可以得出 $\hat{x}_k(3), \hat{x}_k(4), \cdots, \hat{x}_k(m)$ 所满

足的等式．由上述讨论不难看出，我们所导出的这套算法是具有递推特性的．只要知道了 $\hat{x}_k(1), \hat{x}_{k-1}(1), \cdots$ 和 $x_k, x_{k-1}, \cdots$，我们就可以直接应用（3.26）式而算出 $\hat{x}_k(2)$，然后又可以算出 $\hat{x}_k(3)$，如此等等．

显然，博克斯-詹金斯方法的关键在于要有一个 ARMA(n,p) 模型．当然，这最终归结为确定阶数 n, p 和估计系数 $a_1, a_2, \cdots, a_n, c_1, c_2, \cdots, c_p$．关于这些问题，我们已经能够解决了．

以下我们将说明，利用博克斯-詹金斯方法，也可以求解一类非平稳序列的预报问题．

设 $\{x_k\}$ 不是一个平稳序列，但它的 d 阶增量 $\{\Delta^d x_k\}$ 却是一个平稳序列．并满足某个 ARMA(n, p) 模型

$$A(q^{-1})y_k = C(q^{-1})e_k \tag{3.27}$$

其中 $y_k = \Delta^d x_k$，$\Delta = 1 - q^{-1}$，$\{e_k\}$是零均值的白噪声，并且

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_pq^{-p}$$

多项式 $A(z)$ 和 $C(z)$ 的零点皆在单位圆外．

我们也可以把模型(3.27)写成下述形式

$$(1 - q^{-1})^d A(q^{-1})x_k = C(q^{-1})e_k \tag{3.28}$$

或

$$A^*(q^{-1})x_k = C(q^{-1})e_k$$

其中 $A^*(q^{-1}) = (1 - q^{-1})^d A(q^{-1})$．因为 $A^*(z)$ 在单位圆上有 d 重的零点，所以模型（3.28）所确定的序列是非平稳的．博克斯和詹金斯把这种模型叫做自回归积分滑动平均模型．简记作 ARIMA(n, d, p)．

显然，我们可以把 $A^*(q^{-1})$ 写成下述形式

$$A^*(q^{-1}) = 1 + \alpha_1q^{-1} + \alpha_2q^{-2} + \cdots + \alpha_{n+d}q^{-(n+d)}$$

于是模型(3.28)可以写成

$$x_k + \alpha_1x_{k-1} + \cdots + \alpha_{n+d}x_{k-n-d} = e_k + c_1e_{k+1} + \cdots + c_pe_{k-p}$$

当然未知参数 $\alpha_1, \alpha_2, \cdots, \alpha_{n+d}, c_1, c_2, \cdots, c_p$ 是可以估计的．从

而我们可以应用博克斯-詹金斯方法求解关于序列 $\{x_k\}$ 的预报问题. 能够得出形式完全类似于 (3.25) 或 (3.26) 式的递推预报算法.

例如,考虑非平稳序列

$$x_k = e_k + \gamma_{k-1}e_{k-1} + \gamma_{k-2}se_{k-1} + \gamma_{k-3}s^2e_{k-2} \qquad (3.29)$$

的预报问题(参看文献[14]),其中 γ_{k-1}, γ_{k-2}, γ_{k-3} 是常数. $s=(1-q^{-1})^{-1}=\Delta^{-1}$. 不难把模型(3.29)写成下述形式

$$\begin{aligned}\Delta^2x_k &= \Delta^2e_k + \gamma_{k-1}\Delta^2e_{k-1} + \gamma_{k-2}\Delta e_{k-1} + \gamma_{k-3}e_{k-2}\\ &= (e_k - 2e_{k-1} + e_{k-2}) + \gamma_{k-1}(e_{k-1} - 2e_{k-2} + e_{k-3})\\ &\quad + \gamma_{k-2}(e_{k-1} - e_{k-2}) + \gamma_{k-3}e_{k-1}\\ &= e_k + b_1e_{k-1} + b_2e_{k-2} + b_3e_{k-3}\end{aligned}$$

其中

$$b_1 = \gamma_{k-1} + \gamma_{k-2} + \gamma_{k-3} - 2$$
$$b_2 = 1 - 2\gamma_{k-1} - \gamma_{k-2}$$
$$b_3 = \gamma_{k-1}$$

这就是说 x_k 是 ARIMA (0, 2, 3) 过程: $d=2$, $A(z)=1$, $C(z)=1+b_1z+b_2z^2+b_3z^3$

所求的预报公式是

$$\hat{x}_k(1) = \gamma_{k-1}e_k + \gamma_{k-2}\{se_k\} + \gamma_{k-3}\{s^2e_k\}$$

§ 3.5 具有 MA 噪声的系统预报的奥斯特隆姆方法

考虑具有 MA 噪声系统的预报问题

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k \qquad (3.30)$$

其中

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$
$$B(q^{-1}) = b_0 + b_1q^{-1} + \cdots + b_mq^{-m}$$
$$C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_pq^{-p}$$

并且假定 $C(z)$ 的零点皆在单位圆外,即

$$v_k = C(q^{-1})e_k$$

形成一个 MA 噪声，其中 e_k 是零均值的白噪声，y_k 是输出，u_k 是输入。

因为我们关心的是预报问题，控制输入 u_k 的作用可以忽略。故暂且不妨假定 $u_k=0$。所以我们有

$$A(q^{-1})y_k = C(q^{-1})e_k \tag{3.31}$$

奥斯特隆姆 (Åström) 在他的专著[2]中，证明了下述的关于寻求最小平方误差预报算法的定理。

定理 3.7 设 $\{y_k\}$ 是一个随机序列，它满足

$$A(q^{-1})y_k = C(q^{-1})e_k$$

其中 $\{e_k\}$ 是一个零均值的正态白噪声序列，$\mathrm{var}\{e_k\}=\sigma^2$。并且假定 $A(z)$ 和 $C(z)$ 的零点皆在单位圆外，则 y_k 在最小均方误差意义下的向前 h 步的预报 $\hat{y}_k(h)$ 满足

$$\hat{y}_k(h) = G(q^{-1})C^{-1}(q^{-1})y_k$$

其中 $G(q^{-1})$ 满足

$$A^{-1}(q^{-1})C(q^{-1}) = F(q^{-1}) + q^{-h}G(q^{-1})A^{-1}(q^{-1})$$

而 $F(q^{-1})$ 是 q^{-1} 的 $h-1$ 次多项式

$$F(q^{-1}) = 1 + f_1q^{-1} + \cdots + f_{h-1}q^{-(h-1)}$$

预报的误差是

$$\tilde{y}_k(h) = e_{k+h} + f_1e_{k+h-1} + \cdots + f_{h-1}e_{k+1}$$

它的方差是

$$\mathrm{var}\{\tilde{y}_k(h)\} = \sigma^2(1 + f_1^2 + \cdots + f_{h-1}^2)$$

证明 由于我们有

$$A(q^{-1})y_k = C(q^{-1})e_k$$

以及

$$A^{-1}(q^{-1})C(q^{-1}) = F(q^{-1}) + q^{-h}G(q^{-1})A^{-1}(q^{-1})$$

所以

$$\begin{aligned} y_{k+h} &= A^{-1}(q^{-1})C(q^{-1})e_{k+h} \\ &= F(q^{-1})e_{k+h} + G(q^{-1})A^{-1}(q^{-1})e_k \end{aligned}$$

然而

$$e_k = A(q^{-1})C^{-1}(q^{-1})y_k$$

故有

$$y_{k+h}=F(q^{-1})e_{k+h}+G(q^{-1})C^{-1}(q^{-1})y_k \tag{3.32}$$

由于 $C(z)$ 的一切零点皆在单位圆外，所以 $G(q^{-1})C^{-1}(q^{-1})y_k$ 有意义。

对于 $y_k, y_{k-1}, \cdots$ 的任意一个线性函数 $\hat{y}=\hat{y}(y_k, y_{k-1}, \cdots)$，以它为 y_{k+h} 的预报值，则预报误差为

$$\begin{aligned}\tilde{y}_k(h)&=y_{k+h}-\hat{y}\\&=F(q^{-1})e_{k+h}+G(q^{-1})C^{-1}(q^{-1})y_k-\hat{y}\end{aligned}$$

注意到 $\{e_k\}$ 是独立的随机变量序列. 而 $y_k, y_{k-1}, \cdots$ 仅与 e_k, e_{k-1}, $\cdots$有关，所以 $F(q^{-1})e_{k+h}$ 与 y_k, y_{k-1}, $\cdots$独立. 故 $F(q^{-1})e_{k+h}$ 与 $G(q^{-1})C^{-1}(q^{-1})y_k-\hat{y}$ 是独立的. 由此得出

$$\begin{aligned}E\{\tilde{y}_k(h)^2\}&=E\{[y_{k+h}-\hat{y}]^2\}\\&=E\{[F(q^{-1})e_{k+h}]^2\}+E\{[G(q^{-1})C^{-1}(q^{-1})y_k-\hat{y}]^2\}\end{aligned}$$

可见，只要取

$$\hat{y}=G(q^{-1})C^{-1}(q^{-1})y_k$$

就可使预报误差的平方均值 $E\{\tilde{y}_k(h)^2\}$ 达到最小. 所以我们所寻求的预报公式是

$$\hat{y}_k(h)=G(q^{-1})C^{-1}(q^{-1})y_k$$

或者写成

$$C(q^{-1})\hat{y}_k(h)=G(q^{-1})y_k$$

此外由(3.32)式不难得出

$$\begin{aligned}\tilde{y}_k(h)&=F(q^{-1})e_{k+h}\\&=e_{k+h}+f_1e_{k+h-1}+\cdots+f_{h-1}e_{k+1}\end{aligned}$$

注意到 $\{e_k\}$ 是零均值的正态白噪声. 即得

$$\mathrm{var}\{\tilde{y}_k(h)\}=\sigma^2\{1+f_1^2+f_2^2+\cdots+f_{h-1}^2\} \qquad ■$$

对于控制输入 u_k 不为零的情形，完全类似地有：

定理 3.8 设 $\{y_k\}$ 是随机序列，它满足

$$A(q^{-1})y_k=B(q^{-1})u_k+C(q^{-1})e_k$$

$\{e_k\}$ 是零均值的独立随机变量序列，并且 $\mathrm{var}\{e_k\}=\sigma^2$, $\{u_k\}$ 与 $\{e_k\}$独立. 多项式 $A(z)$ 与 $C(z)$ 的零点皆在单位圆外，则 y_k

在最小均方误差意义下的向前 h 步的预报 $\hat{y}_k(h)$ 满足

$$\hat{y}_k(h) = F(q^{-1})B(q^{-1})C(q^{-1})^{-1}u_{k+h} + C(q^{-1})^{-1}G(q^{-1})y_k \tag{3.33}$$

或者

$$C(q^{-1})\hat{y}_k(h) = F(q^{-1})B(q^{-1})u_{k+h} + G(q^{-1})y_k$$

其中 $F(q^{-1})$ 和 $G(q^{-1})$ 满足

$$A(q^{-1})^{-1}C(q^{-1}) = F(q^{-1}) + q^{-h}A(q^{-1})^{-1}G(q^{-1})$$

$F(q^{-1})$ 是 q^{-1} 的 $h-1$ 次多项式

$$F(q^{-1}) = 1 + f_1q^{-1} + f_2q^{-2} + \cdots + f_{h-1}q^{-(h-1)}$$

于是 $G(q^{-1})$ 是 q^{-1} 的 $n-1$ 次多项式

$$G(q^{-1}) = g_0 + g_1q^{-1} + \cdots + g_{n-1}q^{-(n-1)}$$

预报误差是

$$\tilde{y}_k(h) = F(q^{-1})e_{k+h}$$

其方差是

$$\mathrm{var}\{\tilde{y}_k(h)\} = \sigma^2(1 + f_1^2 + \cdots + f_{h-1}^2)$$

这个定理的证明与定理 3.7 类似，故不再重复。

§ 3.6 自校正预报方法

设随机信号$\{y_k\}$满足

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k \tag{3.34}$$

其中 $\{u_k\}$ 是输入信号，$\{e_k\}$ 是零均值的独立的随机变量序列，并且 $E\{e_k^2\} = \sigma^2$. 而

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \cdots + b_mq^{-m}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_pq^{-p}$$

这里，我们假定模型 (3.34) 的参数 $a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, \cdots, c_p$ 是未知的. 用 $\hat{y}_k(h)$ 表示输出信号 y_{k+h} 的基于$\{y_k, y_{k-1}, \cdots\}$的线性最小均方误差估值. 则由(3.33)式知

$$\hat{y}_k(h) = \frac{G(q^{-1})}{C(q^{-1})}y_k + \frac{F(q^{-1})}{C(q^{-1})}B(q^{-1})u_{k+h} \tag{3.35}$$

其中 $F(q^{-1})$, $G(q^{-1})$ 用下述恒等式确定

$$C(q^{-1}) = F(q^{-1})A(q^{-1}) + q^{-h}G(q^{-1}) \tag{3.36}$$

而

$$F(q^{-1}) = 1 + f_1q^{-1} + \cdots + f_{h-1}q^{-(h-1)}$$
$$G(q^{-1}) = g_0 + gq^{-1} + \cdots + g_{n-1}q^{-(n-1)}$$

由于 $a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, \cdots, c_p$ 是未知的，所以参数 $f_1, f_2, \cdots, f_{h-1}, g_0, g_1, \cdots, g_{n-1}$ 也是不能靠它们直接确定的. 一种明显的方法就是首先对模型(3.34)的未知参数进行辨识，然后再用等式 (3.36) 确定 $F(q^{-1})$ 和 $G(q^{-1})$. 从而得出预报公式(3.35). 另一种方法是由维特马克 (B. Wittenmark) 等人根据自校正调节器的思想提出的自校正预报方法[3]. 其特点是不需要去辨识模型(3.34)的未知参数而直接估计预报算法(3.35)式的参数. 为做到这一点，需要把预报公式(3.35)进行适当改变. 令

$$y_k = \hat{y}_{k-h}(h) + \varepsilon_k$$

于是由(3.35)式有

$$y_k = G(q^{-1})y_{k-h} - (C(q^{-1}) - 1)\hat{y}_{k-h}(h) + F(q^{-1})B(q^{-1})u_k + \varepsilon_k$$

如果置

$$A^*(q^{-1}) = G(q^{-1})$$
$$B^*(q^{-1}) = C(q^{-1}) - 1$$
$$C^*(q^{-1}) = F(q^{-1})B(q^{-1})$$

虽然它们的系数是未知的，但阶次却是已知的，故可令

$$A^*(q^{-1}) = \alpha_0 + \alpha_1q^{-1} + \cdots + \alpha_{n-1}q^{-(n-1)}$$
$$B^*(q^{-1}) = \beta_1q^{-1} + \cdots + \beta_pq^{-p}$$
$$C^*(q^{-1}) = \gamma_0 + \gamma_1q^{-1} + \cdots + \gamma_{m+n-1}q^{-(m+h-1)}$$

此时我们有

$$y_k = A^*(q^{-1})y_{k-h} - B^*(q^{-1})\hat{y}_{k-h}(h) + C^*(q^{-1})u_k + \varepsilon_k \tag{3.37}$$

进一步置

$$\phi(k)^\tau = [y_{k-h}, y_{k-h-1}, \cdots, y_{k-h+1}, -\hat{y}_{k-h-1}(h), -\hat{y}_{k-h-2}(h), \cdots$$

$$\cdots \hat{y}_{k-h-p}(h), u_k, u_{k-1}, \cdots, u_{k-m-h+1}]$$

$$\theta^{\tau} = [\alpha_0, \alpha_1, \cdots, \alpha_{n-1}, \beta_1 \cdots \beta_p, \gamma_0, \gamma_1 \cdots, \gamma_{m+h-1}]$$

则(3.37)式可以写成

$$y_k = \phi(k)^{\tau}\theta + \varepsilon_k$$

于是未知参数向量 θ 可由下述递推公式确定

$$\hat{\theta}(k+1) = \hat{\theta}(k) + M(k+1)\{y_{k+1} - \phi(k+1)^{\tau}\hat{\theta}(k)\}$$

$$M(k+1) = \frac{P(k)\phi(k+1)}{\lambda + \phi(k+1)^{\tau}P(k)\phi(k+1)}$$

$$P(k+1) = \frac{1}{\lambda}[I - M(k+1)\phi(k+1)^{\tau}]P(k) \quad (3.38)$$

其中 λ 是遗忘因子，初值 $P(0)$ 和 $\hat{\theta}(0)$ 可适当选取.

上述的一组递推算法， 给出了公式(3.37)中未知参数的相应估值. 在估计 $\hat{\theta}(k)$ 时，我们将利用直到时刻 k 的观测值 y_k. 利用估值 $\hat{\theta}(k)$ 和公式(3.37)，可以得出向前 h 步的预报估值算法.

$$\hat{y}_k(h) = \hat{A}^*(q^{-1})y_k - \hat{B}^*(q^{-1})\hat{y}_k(h) + \hat{C}^*(q^{-1})u_{k+h} \quad (3.39)$$

其中多项式 $\hat{A}^*(q^{-1})$, $\hat{B}^*(q^{-1})$, $\hat{C}^*(q^{-1})$ 的系数是$\hat{\theta}(k)$的相应分量.

由于 $B^*(q^{-1})$ 具有形式

$$B^*(q^{-1}) = \beta_1 q^{-1} + \cdots + \beta_p q^{-p}$$

所以在应用公式(3.39)进行预报时，要用到估值

$$\hat{y}_{k-1}(h), \hat{y}_{k-2}(h), \cdots, \hat{y}_{k-p}(h)$$

而算法(3.38)和(3.39)式都具有递推的特征. 所以在开始计算时，$\hat{y}_{k_0-1}(h), \hat{y}_{k_0-2}(h), \cdots, \hat{y}_{k_0-p}(h)$ 之值可以适当地猜测，k_0 是初始时刻.

上述这种预报算法，不仅适用于模型参数未知的系统，而且由于算法的递推特征，对于系数随时间变化的系统也有一定的适用性，所以它是一种具有自适应性质的预报算法. 按维特马克，把这种算法叫做自校正预报算法.

在实际应用中，u_k 表示预报因子，它是可以事先给定的. 如果所涉及到的预报因子不仅一个而是多个，例如：$u_1(k), u_2(k), \cdots$,

$u_M(k)$，则不难看出，相应的预报算法改变为

$$\hat{y}_k(h) = \hat{A}^*(q^{-1})y_k - \hat{B}^*(q^{-1})\hat{y}_k(h) + \sum_{j=1}^{M} \hat{C}_j^*(q^{-1})u_j(k+h)$$

其中

$$\hat{A}^*(q^{-1}) = \hat{\alpha}_0 + \hat{\alpha}_1 q^{-1} + \cdots + \hat{\alpha}_{n-1} q^{-(n-1)}$$

$$\hat{B}^*(q^{-1}) = \beta_1 q^{-1} + \cdots + \beta_p q^{-p}$$

$$\hat{C}_j^*(q^{-1}) = \hat{\gamma}_{j0} + \hat{\gamma}_{j1} q^{-1} + \cdots + \hat{\gamma}_{j,m+h-1} q^{-(m+h-1)}$$

$$\hat{\theta}(k)^\tau = (\hat{\alpha}_0, \hat{\alpha}_1, \cdots, \hat{\alpha}_{n-1}, \hat{\beta}_1, \cdots, \hat{\beta}_p, \hat{\gamma}_{10}, \cdots, \gamma_{j,m+h-1}, \hat{\gamma}_{20}, \cdots, \hat{\gamma}_{2,m+h-1}, \cdots, \hat{\gamma}_{M,0}, \cdots, \hat{\gamma}_{M,m+h-1})$$

$$\boldsymbol{\phi}(k)^\tau = [y_{k-h}, \cdots, y_{k-h-n+1}, -\hat{y}_{k-h-1}(h), -\hat{y}_{k-h-2}(h), \cdots, -\hat{y}_{k-h-p}(h), u_1(k-1), \cdots, u_1(k-m-h+1), u_2(k-1), \cdots, u_2(k-m-h+1), \cdots, u_M(k-1), \cdots, u_M(k-m-h+1)]$$

而 $\hat{\boldsymbol{\theta}}(k)$ 则由算法(3.38)确定。

§ 3.7 卡尔曼预报方法

我们考虑用状态空间模型描写的系统的预报问题。设

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \mathbf{w}(k) \tag{3.40}$$

$$\mathbf{y}(k) = H(k)\mathbf{x}(k) + \mathbf{v}(k) \tag{3.41}$$

其中 $\mathbf{x}(k)$ 为 n 维的状态向量，$\mathbf{y}(k)$ 为 m 维的观测向量，$\mathbf{w}(k)$ 是 n 维的系统(装置)噪声，$\mathbf{v}(k)$ 是 m 维的观测噪声。假定 $\mathbf{w}(k)$ 和 $\mathbf{v}(k)$ 相互独立，它们都是正态的，并且

$$E\{\mathbf{w}(k)\} = 0; \quad E\{\mathbf{v}(k)\} = 0$$

对任何的 k 和 j 皆有

$$E\{\mathbf{w}(k)\mathbf{w}(j)^\tau\} = Q\delta_{kj} \quad E\{\mathbf{w}(k)\mathbf{v}(j)^\tau\} = 0$$

$$E\{\mathbf{v}(k)\mathbf{v}(j)^\tau\} = R\delta_{kj}$$

初始值 $x(0)$ 也是正态的，并且与 $\mathbf{w}(k)$，$\mathbf{v}(k)$ 相互独立。

我们的问题是基于观测数据 $Y_k = \{\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(k)\}$，求系统状态 $\mathbf{x}(j)(j > k)$ 的在线性最小方差意义下的预报估值

$\hat{\mathbf{x}}(j|k)$.

不难看出，对于用方程(3.40)和(3.41)所描写的系统，下述结论必成立：

1) 随机序列 $\{x(k)\}$ 和 $\{y(k)\}$ 都服从高斯(正态)分布. 如果 $\boldsymbol{\mu}_x(0)=E\{\mathbf{x}(0)\}=0$，则它们的均值皆为零.

2) 对一切 $k\geqslant j(j=0,1,\cdots)$，皆有

$$E\{\mathbf{x}(j)\mathbf{w}(k)^\tau\}=0,\quad E\{\mathbf{y}(j)\mathbf{w}(k)^\tau\}=0$$

3) 对一切 $k=1,2,\cdots$ 和 $j=0,1,2,\cdots$ 皆有

$$E\{\mathbf{x}(j)\mathbf{v}(k)^\tau\}=0$$

4) 对一切 $k>j$，皆有

$$E\{\mathbf{y}(j)\mathbf{v}(k)^\tau\}=0$$

对于预报问题，有下述的定理

定理 3.9 对于系统(3.40)和(3.41)，如果滤波估值 $\hat{\mathbf{x}}(k|k)$ 和滤波误差 $\tilde{\mathbf{x}}(k|k)=\mathbf{x}(k)-\hat{\mathbf{x}}(k|k)$ 的方差矩阵 $P(k|k)$ 对于某个 k 是已知的. 则对于一切的 $j>k$，我们有

1) 线性最小方差预报值被下式确定

$$\hat{\mathbf{x}}(j|k)=\Phi(j,k)\hat{\mathbf{x}}(k|k) \tag{3.42}$$

其中

$$\Phi(j,k)=\Phi(j-1)\Phi(j-2)\cdots\Phi(k) \tag{3.43}$$

2) 由预报误差 $\tilde{\mathbf{x}}(j|k)=\mathbf{x}(j)-\hat{\mathbf{x}}(j|k)$ 所形成的随机序列 $\{\tilde{\mathbf{x}}(j|k),j=k+1,k+2,\cdots\}$ 是零均值的高斯-马尔可夫 Markov 序列. 其方差矩阵满足

$$P(j|k)=\Phi(j,k)P(k|k)\Phi(j,k)^\tau+\sum_{i=k+1}^{j}\Phi(j,i)Q\Phi(j,i)^\tau \tag{3.44}$$

证明 由于在正态的情形，线性最小方差预报就是

$$\hat{\mathbf{x}}(j|k)=E\{\mathbf{x}(j)|\mathbf{y}(1),\cdots,\mathbf{y}(k)\}$$

而且对于 $j\geqslant k+1$，有

$$\mathbf{x}(j)=\Phi(j,k)\mathbf{x}(k)+\sum_{i=k+1}^{j}\Phi(j,i)\mathbf{w}(i-1)$$

所以

$$\hat{\mathbf{x}}(j|k) = E\left\{[\Phi(j,k)\mathbf{x}(k) + \sum_{i=k+1}^{j}\Phi(j,i)\mathbf{w}(i-1)]|\mathbf{y}(1),\cdots,\mathbf{y}(k)\right\}$$

$$= \Phi(j,k)E\{\mathbf{x}(k)|\mathbf{y}(1),\cdots,\mathbf{y}(k)\} + \sum_{i=k+1}^{j}\Phi(j,i)E\{\mathbf{w}(i-1)|\mathbf{y}(1),\cdots,\mathbf{y}(k)\} \tag{3.45}$$

由前面的结论 2)，对于一切 $j \geqslant k+1$，随机向量 $\{\mathbf{w}(i-1), i=k+1,\cdots,j\}$ 和 $\{\mathbf{y}(1),\cdots,\mathbf{y}(k)\}$ 是不相关的,但它们都具有正态分布,所以也是独立的. 从而对于 $i=k+1,\cdots,j$ 有

$$E\{\mathbf{w}(i-1)|\mathbf{y}(1),\cdots,\mathbf{y}(k)\} = E\{\mathbf{w}(i-1)\} = 0$$

于是(3.45)式变成

$$\hat{\mathbf{x}}(j|k) = \Phi(j,k)\hat{\mathbf{x}}(k|k) \qquad j > k$$

这就得出了(3.42)式.

进一步对于给定的 k 和一切 $j > k$，我们有

$$\tilde{\mathbf{x}}(j|k) = \mathbf{x}(j) - \hat{\mathbf{x}}(j|k)$$

$$= \Phi(j,k)\mathbf{x}(k) + \sum_{i=k+1}^{j}\Phi(j,i)w(i-1) - \Phi(j,k)\hat{\mathbf{x}}(k|k)$$

$$= \Phi(j,k)\tilde{\mathbf{x}}(k|k) + \sum_{i=k+1}^{j}\Phi(j,i)w(i-1) \tag{3.46}$$

由定理的假设与卡尔曼滤波的性质知 $\hat{\mathbf{x}}(k|k)$ 是零均值的高斯变量,所以 $\tilde{\mathbf{x}}(k|k)$ 也是零均值的高斯变量. 于是由(3.46)式可见, $\{\tilde{\mathbf{x}}(j|k), j=k+1, k+2,\cdots\}$ 必为零均值的高斯序列. 以下来说明，这一序列的马尔可夫性. 用K表示集合 $\{k+1, k+2,\cdots\}$，设 $t_1 < t_2 < \cdots < t_{q-1} < t_q$ 是K中的任意 q 个数. 不妨记 $t_q = q$，$t_{q-1} = q-1$，则

$$\tilde{\mathbf{x}}(q|k) = \Phi(q,k)\tilde{\mathbf{x}}(k|k) + \sum_{i=k+1}^{q}\Phi(q,i)\mathbf{w}(i-1)$$

$$
\begin{aligned}
&= \Phi(q-1)\Phi(q-1,k)\tilde{\mathbf{x}}(k|k) + \Phi(q,q)\mathbf{w}(q-1) \\
&\quad + \sum_{i=k+1}^{q-1} \Phi(q,i)\mathbf{w}(i-1) \\
&= \Phi(q-1)\Phi(q-1,k)\tilde{\mathbf{x}}(k|k) \\
&\quad + \Phi(q-1)\sum_{i=k+1}^{q-1} \Phi(q-1,i)\mathbf{w}(i-1) + \mathbf{w}(q-1) \\
&= \Phi(q-1)\Big[\Phi(q-1,k)\tilde{\mathbf{x}}(k|k) \\
&\quad + \sum_{i=k+1}^{q-1} \Phi(q-1,i)\mathbf{w}(i-1)\Big] \\
&\quad + \mathbf{w}(q-1)
\end{aligned}
$$

注意到(3.46)式后，可由上式得出

$$
\tilde{\mathbf{x}}(q|k) = \Phi(q-1)\tilde{\mathbf{x}}(q-1|k) + \mathbf{w}(q-1) \tag{3.47}
$$

由此式即可得出序列 $\{\tilde{\mathbf{x}}(j|k),\ j = k+1,\ k+2,\ \cdots\}$ 的马尔可夫性。

最后推导公式 (3.44). 注意到

$$
P(j|k) = E\{\tilde{\mathbf{x}}(j|k)\tilde{\mathbf{x}}(j|k)^{\tau}\} \tag{3.48}
$$

以及对一切 $i = k+1,\ k+2,\cdots,j$,

$$
E\{\hat{\mathbf{x}}(k|k)\mathbf{w}(i-1)^{\tau}\} = 0
$$

$$
E\{\mathbf{w}(j)\mathbf{w}(k)^{\tau}\} = 0 \quad j \neq k
$$

把(3.46)式代入(3.48)式中经简化即可得出

$$
\begin{aligned}
P(j|k) &= \Phi(j,k)P(k|k)\Phi(j,k)^{\tau} \\
&\quad + \sum_{i=k+1}^{j} \Phi(j,i)Q\Phi(j,i)^{\tau}
\end{aligned}
$$

■

下面的例子可以说明，用卡尔曼方法寻求以状态方程描写的系统的预报问题的解是有实际意义的。

例：导弹位置和速度的预报问题

设对导弹的位置的观测方程是

$$
\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{w}(k) \tag{3.49}
$$

观测噪声 $\mathbf{w}(k)$ 满足

$$E\{\mathbf{w}(k)\}=0 \quad E\{\mathbf{w}(j)\mathbf{w}(k)\}=\sigma_w^2\delta_{j,k}$$

$\mathbf{x}(k)$ 表示导弹的位置.

如果用 $\mathbf{v}(k)$ 表示导弹的速度，$\mathbf{a}(k)$ 表示导弹的加速度，T 表示采样间隔时间. 则有

$$\mathbf{x}(k+1)=\mathbf{x}(k)+T\mathbf{v}(k)+\frac{1}{2}T^2\mathbf{a}(k) \tag{3.50}$$

$$\mathbf{v}(k+1)=\mathbf{v}(k)+T\mathbf{a}(k) \tag{3.51}$$

假定加速度 $\mathbf{a}(k)$ 是随机的,它是白噪声,并且满足

$$E\{\mathbf{a}(k)\}=0 \quad E\{\mathbf{a}(j)\mathbf{a}(k)\}=\sigma_a^2\delta_{j,k}$$

引入状态变量

$$\mathbf{x}_k=\begin{pmatrix}x(k)\\ v(k)\end{pmatrix}$$

置

$$\Phi=\begin{pmatrix}1 & T\\ 0 & 1\end{pmatrix} \qquad \Gamma=\begin{pmatrix}\frac{1}{2}T^2\\ T\end{pmatrix}$$

$$H=(1,0)$$

则得系统的状态方程

$$\mathbf{x}_{k+1}=\Phi x_k+\Gamma\mathbf{a}(k) \tag{3.52}$$

和观测方程

$$\mathbf{y}(k)=H\mathbf{x}_k+\mathbf{w}(k) \tag{3.53}$$

导弹位置和速度的预报问题是：基于所观测的导弹的位置 $\{\mathbf{y}(1),\cdots,\mathbf{y}(k)\}$，寻求导弹在将来时刻 $k+h$ 时的位置 $\mathbf{x}(k+h)$ 和速度 $\mathbf{v}(k+h)$ 的预报值 $\hat{\mathbf{x}}(k+h|k)$ 和 $\mathbf{v}(k+h|k)$. 显然这个问题可依据模型(3.52)和(3.53)用卡尔曼方法来解决.

第四章　动态系统时变参数的辨识

本章内容包括两部分：从解决动态系统预报问题的特殊观点出发来讨论时变参数的辨识问题和从一般观点出发来讨论时变参数的辨识问题．在前一种观点下，我们引出了一种“权宜”的算法，它对于解决预报问题是很有效的．在后一种观点下，我们应用多项式逼近的方法引出了关于时变参数辨识的一种途径．

§4.1　预报模型的参数辨识准则

通常，解决动态系统的预报问题要从资料的收集和分析着手．在此基础上通过适当地应用辨识方法，建立起系统的数学模型．对于控制系统的设计，一般也要进行系统辨识．但预报问题与控制问题的本质存在着不同之处．虽然，系统的控制模型经过适当地处理后，可以用来作为预报模型，但是专门为解决预报问题所建立起的模型，对于解决预报问题而言，可能有更多的长处．因此有必要发展一套适用于预报问题的辨识方法．

我们首先来讨论关于辨识准则的确定问题．不失一般性，考虑如下形式的向前一步的预报误差模型：

$$\mathbf{y}(k) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + \mathbf{e}(k) \tag{4.1}$$

其中 $\mathbf{y}(k)$ 是输出，$\mathbf{u}(k)$ 是输入，$\mathbf{e}(k)$ 是随机噪声，假定它与 $f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k]$ 不相关．而

$$\mathbf{Y}_{k-1} = \{\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(k-1)\}$$
$$\mathbf{U}_k = \{\mathbf{u}(0), \mathbf{u}(1), \cdots, \mathbf{u}(k)\}$$

$\theta(k)$ 是 m 维的参数．

不难看出，对于模型(4.1)而言，如果已知资料 $\mathbf{Y}_{k-1}$，$\mathbf{U}_k$，和参量 $\theta(k)$ 的值 θ，则系统输出的向前一步的预报公式是

$$\hat{\mathbf{y}}(k|\theta) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta, k] \tag{4.2}$$

相应的预报残差是

$$\begin{aligned}\boldsymbol{\varepsilon}(k, \theta) &= \mathbf{y}(k) - \hat{\mathbf{y}}(k|\theta) \\ &= \mathbf{y}(k) - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta, k]\end{aligned}$$

在 $f[\cdot,\cdot,\theta(k),k]$ 的结构已知的情况下，预报公式(4.2)将依赖于 $\theta(k)$ 的估值．我们自然希望预报的"误差"尽可能的小，也即希望 $\|\boldsymbol{\varepsilon}(k,\theta)\|$ 尽可能的小，而 θ 值的不同将导致残差 $\boldsymbol{\varepsilon}(k,\theta)$ 的不同，所以

$$J_k(\theta) = \|\boldsymbol{\varepsilon}(k,\theta)\|$$

是 θ 的函数．

由于 $\mathbf{y}(k)$ 是被预报量，它是未知的．从而仅可能有资料 $\mathbf{Y}_{k-1}$ 和 $\mathbf{U}_k$．所以应用一般的估值算法仅能得到 $\theta(k-1)$ 的估值 $\hat{\theta}(k-1)$，于是，一般的所谓自适应预报算法，对于模型 (4.1) 而言，仅能具有形式

$$\hat{\mathbf{y}}(k|\hat{\theta}(k-1)) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k-1), k] \tag{4.3}$$

如果系统的参数是快时变的，即 $\theta(k)$ 与 $\theta(k-1)$ 有较大差别．则上述的算法(4.3)将产生较大的预报误差．为减少这一误差，我们应设法得到 $\theta(k)$ 的某种估值 $\hat{\theta}^*(k)$． 进而把预报算法写成

$$\hat{\mathbf{y}}(k|\hat{\theta}^*(k)) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}^*(k), k] \tag{4.4}$$

并称之为含有参数预报值的预报公式．此时预报的残差为

$$\boldsymbol{\varepsilon}(k|\hat{\theta}^*(k)) = \mathbf{y}(k) - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}^*(k), k]$$

以下为方便计，我们称

$$\boldsymbol{\varepsilon}(k|\hat{\theta}(k)) = \mathbf{y}(k) - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k), k]$$

为后验残差，其中 $\hat{\theta}(k)$ 是 $\theta(k)$ 的依赖于资料 $\mathbf{Y}_k$，$\mathbf{U}_k$ 的估值．至此，我们可有如下定义：

定义 设 $\{A_\delta: \delta \in \Delta\}$ 是一族含有参量 δ 的关于动态系统参量 $\theta(k)$ 的估值算法． 如果对于任何的 $\varepsilon > 0$，皆有 $\delta^0 \in \Delta$ 和 $N > 0$，使得当 $k \geqslant N$ 时，由算法 A_{δ^0} 所得出的参量估值 $\hat{\theta}(k)$ 恒满足

$$\|\boldsymbol{\varepsilon}(k, \hat{\theta}(k))\| < \varepsilon \quad \text{a.s}$$

我们就说这族算法满足后验残差一致小准则.

这就是我们所提出的关于预报模型的参数辨识的辨识准则. 下面说明: 我们所熟知的五种递推算法 (RLS, RIV, RGLS, RELS, RML) 通常是不满足后验残差一致小准则的. 如我们所知的,若系统的模型具有如下的形式

$$y_k = \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta} + e_k$$

其中 y_k 是输出, e_k 是白噪声, θ 是未知参数向量, $\boldsymbol{\phi}(k)$ 是由 $y_{k-h}(1 \leqslant h \leqslant n)$, $u(k-j)(0 \leqslant j \leqslant n)$ 及噪声 $e_{k-h}(1 \leqslant h \leqslant n)$ 或它们的某种估值组成的向量, $u(k)$ 是输入. 则可以把上述五种递推算法写成如下的统一的形式

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{P(k-1)\mathbf{z}(k)}{\lambda_k + \boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)} \times \{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\} \tag{4.5}$$

$$P(k) = \frac{1}{\lambda_k}\left[I - \frac{P(k-1)\mathbf{z}(k)}{\lambda_k + \boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)}\boldsymbol{\phi}(k)^\tau\right]P(k-1) \tag{4.6}$$

其中 $\hat{\boldsymbol{\theta}}(k)$ 表示 $\boldsymbol{\theta}$ 的第 k 次 (k 时刻) 估值, $\mathbf{z}(k)$ 是适当确定的向量,它随算法的不同而不同(参看文献[10]); λ_k 是遗忘因子. 与上述统一算法相应的后验残差是

$$\begin{aligned}\boldsymbol{\varepsilon}(k, \hat{\boldsymbol{\theta}}(k)) &= y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k) \\ &= y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1) - \frac{\boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)}{\lambda_k + \boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)} \\ &\quad \times \{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\} \\ &= \left\{1 - \frac{\boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)}{\lambda_k + \boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)}\right\}\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\}\end{aligned}$$

由于 y_k, $\boldsymbol{\phi}(k)$, $\hat{\boldsymbol{\theta}}(k-1)$ 都是已知的, 故 $y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)$ 已被完全确定. 可见,欲使 $|\boldsymbol{\varepsilon}(k, \hat{\boldsymbol{\theta}}(k))|$ 充分小, 就必须使

$$1 - \frac{\boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)}{\lambda_k + \boldsymbol{\phi}(k)^\tau P(k-1)\mathbf{z}(k)}$$

足够小,这就要求 λ_k 充分地接近于零. 而这是公式(4.6)在一般

情况下所不能允许的. 所以，熟知的五种递推算法通常是不能满足后验残差一致小准则的.

§ 4.2 预报模型的参数估计算法

不失一般性,以下仅考虑时变参数的单输出系统

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + e_k \tag{4.7}$$

其中 $\theta(k)$ 是 m 维的时变参数，y_k 是一维的输出，e_k 是随机噪声.

系统(4.7)的一个特殊情形是

$$y_k = \phi(k)^\tau \theta(k) + e_k \tag{4.8}$$

对于这个系统,如果 $\phi(k)$ 可由观测完全确定，e_k 是白噪声，参数 $\theta(k)$ 是非时变的或慢时变的，则其估值算法可写成如下形式

$$\hat{\theta}(k) = \hat{\theta}(k-1) + r_k \mathbf{R}(k)^{-1}\phi(k)\{y_k - \phi(k)^\tau \hat{\theta}(k-1)\} \tag{4.9}$$

$$R(k) = R(k-1) + r_k\{\phi(k)\phi(k)^\tau - R(k-1)\} \tag{4.10}$$

如果 $R(k)$ 可以被直接确定，则(4.10)式能够省掉. 可见递推算法的主要部分是(4.9)式. 对于这一算法,其后验残差是

$$\begin{aligned}\epsilon(k, \hat{\theta}(k)) &= y_k - \phi(k)^\tau \hat{\theta}(k)\\ &= y_k - \phi(k)^\tau \hat{\theta}(k-1) - r_k \phi(k)^\tau R(k)^{-1}\phi(k)\\ &\quad \times\{y_k - \phi(k)^\tau \hat{\theta}(k-1)\}\\ &= (1 - r_k\phi(k)^\tau R(k)^{-1}\phi(k))\{y_k - \phi(k)^\tau \hat{\theta}(k-1)\}\end{aligned}$$

由此可见,只要置

$$R(k) = \begin{pmatrix} \|\phi(k)\|^2 & & & 0 \\ & \|\phi(k)\|^2 & & \\ & & \ddots & \\ 0 & & & \|\phi(k)\|^2 \end{pmatrix}$$

就有

$$\epsilon(k, \hat{\theta}(k)) = (1 - r_k)\{y_k - \phi(k)^\tau \hat{\theta}(k-1)\}$$

于是只要使 r_k 充分接近于 1，就可使后验残差 $\epsilon(k, \hat{\theta}(k))$ 任

意小．从而我们得出了如下的估值算法

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{r_k}{\|\phi(k)\|^2}\phi(k)\{y_k-\phi(k)^\tau\hat{\theta}(k-1)\}$$

而对于时变参数的情形，算法则改变成

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{\delta}{\|\phi(k)\|^2}\phi(k)\{y_k-\phi(k)^\tau\hat{\theta}(k-1)\} \tag{4.11}$$

其中 δ 是适当的常数．显然当 $\delta=1$ 时，有 $\varepsilon(k,\hat{\theta}(k))=0$．

对于系统(4.8)，如果 $\phi(k)$ 仅由 $\mathbf{Y}_{k-1},\mathbf{U}_k$ 决定，则当参数值取成 $\hat{\theta}(k-1)$ 时，向前一步的预报公式是

$$\hat{y}(k|\hat{\theta}(k-1))=\phi(k)^\tau\hat{\theta}(k-1)$$

从而有

$$\frac{\partial}{\partial_{\theta(k-1)}}\hat{y}(k|\hat{\theta}(k-1))=\phi(k)$$

故算法(4.11)又可改写为

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{\delta}{\left\|\frac{\partial}{\partial_{\theta(k-1)}}\hat{y}(k|\hat{\theta}(k-1))\right\|^2}$$

$$\cdot\frac{\partial}{\partial_{\theta(k-1)}}\hat{y}(k|\hat{\theta}(k-1))\{y_k-\phi(k)^\tau\hat{\theta}(k-1)\}$$

类似地对系统(4.7)有

$$\hat{y}(k|\hat{\theta}(k-1))=f[\mathbf{Y}_{k-1},\mathbf{U}_k,\hat{\theta}(k-1),k]$$

$$\nabla_{\theta(k-1)}f[k,\hat{\theta}(k-1)]=\frac{\partial}{\partial_{\theta(k-1)}}f[\mathbf{Y}_{k-1},\mathbf{U}_k,\hat{\theta}(k-1),k]$$

$$=\frac{\partial}{\partial_{\theta(k-1)}}\hat{y}(k|\hat{\theta}(k-1))$$

所以平行于算法(4.11)，可以把关于系统(4.7)的时变参数 $\theta(k)$ 的估值算法写成

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{\delta}{\|\nabla_{\theta(k-1)}f[k,\hat{\theta}(k-1)]\|^2}$$

$$\cdot \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]$$
$$\{\mathbf{Y}_k - f[y_{k-1}, \mathbf{U}_k, \hat{\theta}(k-1), k]\} \tag{4.12}$$

其中 δ 是适当确定的常数.

可以说明,在适当的条件下,算法(4.12)是满足后验残差一致小准则的. 事实上,我们有如下的定理:

定理 4.1 对于系统

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + e_k$$

无论其噪声统计特性如何,如果有常数 $c > 0$,使得对一切 k 皆有

$$|y_k| < c \quad \text{a.s}$$

对一切 k, 一切 y_k 和 $u(k)$, 一切 $\beta \in R^m$, 皆有

$$|f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \beta, k]| < c \quad \text{a.s}$$

设 $\{\hat{\theta}(k)\}$ 是由算法(4.12)所确定的,并对 $\delta > 0$ 一致地有

$$\lim_{k\to\infty} \frac{\nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^\tau \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^2} = \mu > 0 \quad \text{a.s} \tag{4.13}$$

其中 $\overline{\theta(k)}$ 由下式决定

$$f[k, \hat{\theta}(k)] = f[k, \hat{\theta}(k-1)] + \nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^\tau \tilde{\theta}(k)$$

此处

$$\tilde{\theta}(k) = \hat{\theta}(k) - \hat{\theta}(k-1)$$
$$f[k, \theta] = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta, k]$$

则对于任何的 $\varepsilon > 0$, 必有 $\delta > 0$ 和 $N > 0$, 使与之相应的 $\{\hat{\theta}(k)\}$所导出的后验残差,当 $k \geqslant N$时,恒满足

$$|\varepsilon(k, \hat{\theta}(k))| = |y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k), k]| < \varepsilon \text{ a.s}$$

证明 对 $f[k, \hat{\theta}(k)] = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k), k]$ 在 $\theta(k-1)$ 处应用微分中值公式,有

$$f[k, \hat{\theta}(k)] = f[k, \hat{\theta}(k-1)] + \nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^\tau \tilde{\theta}(k)$$

但是

$$\tilde{\theta}(k) = \hat{\theta}(k) - \hat{\theta}(k-1) = \frac{\delta}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^2}$$
$$\times \nabla_{\theta(k-1)} f[k, \theta(k-1)]\{y_k - f[k, \theta(k-1)]\}$$

所以

$$f[k, \hat{\theta}(k)] = f[k, \hat{\theta}(k-1)] + \delta \frac{\nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^{\tau} \nabla_{\hat{\theta}(k-1)} f[k, \hat{\theta}(k-1)]}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^2} \times \{y_k - f[k, \hat{\theta}(k-1)]\}$$

于是

$$y_k - f[k, \hat{\theta}(k)] = \left(1 - \delta \frac{\nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^{\tau} \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^2}\right) \times \{y_k - f[k, \hat{\theta}(k-1)]\}$$

由于我们有条件(4.13)式,所以对任何 $\varepsilon > 0$, 必有 $N > 0$,使当 $k \geqslant N$时,恒有

$$\mu - \frac{\varepsilon}{6c}\mu \leqslant \frac{\nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^{\tau} \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^2} \leqslant \mu + \frac{\varepsilon}{6c}\mu$$

注意到

$$\frac{\mu\left(1 - \frac{\varepsilon}{6c}\right)}{\mu\left(1 + \frac{\varepsilon}{6c}\right)} = 1 - \frac{\frac{\varepsilon}{3c}}{1 + \frac{\varepsilon}{6c}} > 1 - \frac{\varepsilon}{3c} > 1 - \frac{\varepsilon}{2c}$$

即

$$\frac{1}{\mu\left(1 + \frac{\varepsilon}{6c}\right)} > \frac{1 - \frac{\varepsilon}{2c}}{\mu\left(1 - \frac{\varepsilon}{6c}\right)}$$

由此可见,只要使 δ 满足

$$\frac{1}{\mu\left(1 + \frac{\varepsilon}{6c}\right)} \geqslant \delta \geqslant \frac{1 - \frac{\varepsilon}{2c}}{\mu\left(1 - \frac{\varepsilon}{6c}\right)}$$

当 $k \geqslant N$ 时就恒有

$$1-\delta \frac{\nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^{\tau} \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^{2}}>0$$

以及

$$\left|y_{k}-f[k, \hat{\theta}(k)]\right| \leqslant\left[1-\delta\left(1-\frac{\varepsilon}{6 c}\right) \mu\right]\left|y_{k}-f[k, \hat{\theta}(k-1)]\right|$$

$$\leqslant\left[1-\delta\left(1-\frac{\varepsilon}{6 c}\right) \mu\right]\{|y_{k}|+|f[k, \hat{\theta}(k-1)]|\}$$

$$\leqslant\left[1-\left(1-\frac{\varepsilon}{2 c}\right)\right](c+c)=\varepsilon \quad \text{a.s} \qquad \blacksquare$$

关于条件(4.13)式的注释．这个条件无疑对所讨论的问题是给了一定的限制，但这种限制并不是过苛的．例如对于形如

$$y_{k}=\boldsymbol{\phi}(k)^{\tau} \theta(k)+e_{k}$$

的系统，恒有

$$f[k, \hat{\theta}(k)]-f[k, \hat{\theta}(k-1)]=\boldsymbol{\phi}(k)^{\tau} \tilde{\theta}(k)$$

从而

$$\lim_{k \to \infty} \frac{\nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^{\tau} \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^{2}}$$

$$=\lim_{k \to \infty} \frac{\boldsymbol{\phi}(k)^{\tau} \boldsymbol{\phi}(k)}{\|\boldsymbol{\phi}(k)\|^{2}}=1$$

即条件(4.13)式必然满足．

§4.3 多输出系统的参数估计

在前节，我们仅考虑了单输出系统的参数估计问题．现在来考虑多输出系统的情形．设有多输出预报误差模型所描写的系统

$$\mathbf{y}(k)=f[\mathbf{Y}_{k-1}, \mathbf{U}_{k}, \theta(k), k]+\mathbf{e}(k) \qquad (4.14)$$

其中 $\mathbf{y}(k)$ 是 n 维输出，$\mathbf{u}(k)$ 是 p 维输入，$\theta(k)$ 是 m 维的参量，$\mathbf{e}(k)$ 是 n 维的噪声．而

$$\mathbf{Y}_{k-1} = \{\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(k-1)\}$$
$$\mathbf{U}_k = \{\mathbf{u}(0), \mathbf{u}(1), \cdots, \mathbf{u}(k)\}$$

模型(4.14)的一个特殊的情形是

$$\mathbf{y}(k) = \boldsymbol{\Phi}(k)\boldsymbol{\theta}(k) + \mathbf{e}(k) \tag{4.15}$$

其中 $\boldsymbol{\Phi}(k)$ 是 $\mathbf{Y}_{k-1}$ 和 $\mathbf{U}_k$ 的矩阵值函数。

下面说明，前节所得出的关于单输出系统的时变参数的估值算法，对于多输出系统(4.14)的推广是

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \delta A(k)^{-1}\nabla_{\theta(k-1)}f[k, \boldsymbol{\theta}(k-1)] \times \{y(k) - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\theta}(k-1), k]\} \tag{4.16}$$

其中 $\nabla_{\theta(k-1)}f[k, \boldsymbol{\theta}(k-1)]$ 是 $m \times n$ 的矩阵，它被定义为

$$\nabla_{\theta(k-1)}f[k, \hat{\boldsymbol{\theta}}(k-1)] = \frac{\partial}{\partial\theta} f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta, k]_{\theta=\theta(k-1)}$$

δ 是适当选取的常数．而

$$A(k) = B_k + \frac{1}{a_{r_k}} \Phi_k(B_k)$$

$$B_k = \nabla_{\theta(k-1)}f[k, \hat{\theta}(k-1)] \cdot \nabla_{\theta(k-1)}f[k, \boldsymbol{\theta}(k-1)]^\tau$$

$\Phi_k(B_k)$ 和 a_{r_k} 是通过下述步骤确定的：设 $\operatorname{rank} B_k = r_k$. 用 $\varphi_k(\lambda)$ 表示 B_k 的特征多项式，即 $\varphi_k(\lambda) = \det(\lambda I - B_k)$，置

$$\Phi_k(\lambda) = \frac{\varphi_k(\lambda)}{\lambda^{m-r_k}}$$

$$a_{r_k} = \Phi_k(0)$$

显然，$a_{r_k} \neq 0$，而 $\Phi_k(\lambda)$ 具有形式

$$\Phi_k(\lambda) = \lambda^{r_k} + a_1\lambda^{r_k-1} + \cdots + a_{r_k-1}\lambda + a_{r_k}$$

于是

$$\Phi_k(B_k) = B_k^{r_k} + a_1 B_k^{r_k-1} + \cdots + a_{r_k-1}B_k + a_{r_k}I$$

引理 4.2 $A(k) = B_k + \dfrac{1}{a_{r_k}} \Phi_k(B_k)$ 是一个正定的对称矩阵。

证明 把 $r_k = \operatorname{rank} B$ 简记为 r. 设 $\lambda_1, \lambda_2, \cdots, \lambda_r$ 是 B_k

的 r 个非零特征值。由于 B_k 是非负定的对称矩阵，所以必有正交矩阵 T，使得

$$TB_kT^\tau = D_k$$

此处 D_k 是一个对角形矩阵，不失一般性可以假定 D_k 具有形式

$$D_k = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

另一方面，由于

$$\Phi_k(\lambda_1) = \Phi_k(\lambda_2) = \cdots = \Phi_k(\lambda_r) = 0$$

所以

$$\begin{aligned} T\Phi_k(B_k)T^\tau &= TB_k^rT^\tau + a_1TB_k^{r-1}T^\tau + \cdots + a_{r-1}TB_kT^\tau \\ &\quad + a_rTIT^\tau \\ &= (TB_kT^\tau)^r + a_1(TB_kT^\tau)^{r-1} + \cdots + a_rI \\ &= D_k^r + a_1D_k^{r-1} + \cdots + a_{r-1}D + a_rI \\ &= \begin{pmatrix} \Phi_k(\lambda_1) & & & & & \\ & \Phi_k(\lambda_2) & & & & \\ & & \ddots & & & \\ & & & \Phi_k(\lambda_r) & & \\ & & & & a_r & \\ & & & & & \ddots \\ & & & & & & a_r \end{pmatrix} \\ &= \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & a_r & & \\ & & & & \ddots & \\ & & & & & a_r \end{pmatrix} \end{aligned}$$

于是有

$$T\left(B_k+\frac{1}{a_r}\Phi_k(B_k)\right)T^{\tau}=\begin{pmatrix}\lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_r & & & \\ & & & 1 & & \\ & 0 & & & \ddots & \\ & & & & & 1\end{pmatrix}$$

所以

$$B_k+\frac{1}{a_r}\Phi_k(B_k)=T^{\tau}\begin{pmatrix}\lambda_1 & & & & & & \\ & \lambda_2 & & & 0 & & \\ & & \ddots & & & & \\ & & & \lambda_r & & & \\ & & & & 1 & & \\ & 0 & & & & \ddots & \\ & & & & & & 1\end{pmatrix}T$$

这就证明了 $A(K)$ 是一个对称的正定矩阵. ■

上述引理说明,我们所提出的递推算法(4.16)是有意义的. 进一步可以证明下述定理:

定理 4.3 如果在动态系统

$$y_k=f[\mathbf{Y}_{k-1},\mathbf{U}_k,\theta_k(k),k]+e_k$$

中,输出 y_k 是一维的,即 $f[\mathbf{Y}_{k-1},\mathbf{U}_k,\theta(k),k]$ 是数值函数,则有

$$A(k)^{-1}\nabla_{\theta(k-1)}f[k,\hat{\theta}(k-1)]$$
$$=\frac{1}{\|\nabla_{\theta(k-1)}f[k,\hat{\theta}(k-1)]\|^2}\nabla_{\theta(k-1)}f[k,\hat{\theta}(k-1)] \quad (4.17)$$

即矩阵

$$A(k)^{-1}=\left[B_k+\frac{1}{a_{rk}}\Phi_k(B_k)\right]^{-1}$$

具有非零特征值

$$\frac{1}{\|\nabla_{\theta(k-1)}f[k,\hat{\theta}(k-1)]\|^2}$$

其所对应的特征向量是 $\nabla_{\theta(k-1)}f[k,\hat{\theta}(k-1)]$.

证明 为记号简便计，令

$$\nabla_{\hat{\theta}(k-1)} f[k,\hat{\theta}(k-1)] = \begin{pmatrix} \alpha_1(k) \\ \alpha_2(k) \\ \vdots \\ \alpha_m(k) \end{pmatrix}$$

于是

$$B_k = \begin{pmatrix} \alpha_1(k) \\ \alpha_2(k) \\ \vdots \\ \alpha_m(k) \end{pmatrix} (\alpha_1(k) \ \ \alpha_2(k) \ \ \cdots \ \ \alpha_m(k))$$

显然，rank $B_k = 1$. 而且

$$\det(\lambda 1 - B_k)$$

$$= \begin{bmatrix} \lambda - a_1^2(k) & -\alpha_1(k)\alpha_2(k) & \cdots -\alpha_1(k)\alpha_m(k) \\ -\alpha_1(k)\alpha_2(k) & \lambda - \alpha_2(k)^2 & \cdots -\alpha_2(k)\alpha_m(k) \\ \vdots & \vdots & \vdots \\ -\alpha_1(k)\alpha_m(k) & -\alpha_2(k)\alpha_m(k) & \cdots \lambda - \alpha_m(k)^2 \end{bmatrix}$$

$$= \left(\lambda - \sum_{i=1}^{m} a_i(k)^2\right)\lambda^{m-1}$$

所以 B_k 唯一的非零特征值是 $a = \sum_{i=1}^{m} \alpha_i(k)^2$ 而且

$$\Phi_k(\lambda) = \lambda - a$$

要证明的是

$$\left(B_k + \frac{1}{a}\Phi_k(B_k)\right)^{-1} \begin{pmatrix} \alpha_1(k) \\ \alpha_2(k) \\ \vdots \\ \alpha_m(k) \end{pmatrix} = \frac{1}{a} \begin{pmatrix} \alpha_1(k) \\ \alpha_2(k) \\ \vdots \\ \alpha_m(k) \end{pmatrix}$$

为此只需证

$$\left(B_k + \frac{1}{a}\Phi_k(B_k)\right) \begin{pmatrix} \alpha_1(k) \\ \alpha_2(k) \\ \vdots \\ \alpha_m(k) \end{pmatrix} = a \begin{pmatrix} \alpha_1(k) \\ \alpha_2(k) \\ \vdots \\ \alpha_m(k) \end{pmatrix}$$

不难看出

$$B_k\begin{pmatrix}\alpha_1(k)\\ \alpha_2(k)\\ \vdots\\ \alpha_m(k)\end{pmatrix}=\begin{pmatrix}\alpha_1(k)\\ \alpha_2(k)\\ \vdots\\ \alpha_m(k)\end{pmatrix}(\alpha_1(k)\ \ \alpha_2(k)\ \ \cdots\ \ \alpha_m(k))$$

$$\cdot\begin{pmatrix}\alpha_1(k)\\ \alpha_2(k)\\ \vdots\\ \alpha_m(k)\end{pmatrix}=\sum_{i=1}^{m}\alpha_i(k)^2\begin{pmatrix}\alpha_1(k)\\ \alpha_2(k)\\ \vdots\\ \alpha_m(k)\end{pmatrix}\tag{4.18}$$

所以只需证

$$\Phi_k(B_k)\begin{pmatrix}\alpha_1(k)\\ \alpha_2(k)\\ \vdots\\ \alpha_m(k)\end{pmatrix}=\begin{pmatrix}0\\ 0\\ \vdots\\ 0\end{pmatrix}\tag{4.19}$$

但我们有

$$\Phi_k(B_k)=B_k-aI$$

并注意到(4.18)式，立即可得出(4.19)式. ■

§4.4 多输出系统的分解

这里，我们将说明，从预报模型的参数辨识观点来考虑问题，在一定条件下，可以把多输出系统分解成为在一定意义下与之等效的一些广义单输出系统. 因为从前面的一系列讨论可以看出，我们所提出的参数估值算法，在多输出系统的情形，比单输出系统的情形要复杂得多. 所以，这一节所得的结论是很有意义的.

定义 我们说多输出系统 S 的模型:

$$\mathbf{y}(k)=f[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}(k),k]+\mathbf{v}(k)$$

具有可分离参数的形式，或简称为参数可分离的，如果 $f[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}(k),k]$ 具有形式

$$f[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}(k),k]=\begin{pmatrix}f_1[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}_1(k),k]\\ f_2[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}_2(k),k]\\ \vdots\\ f_n[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}_n(k),k]\end{pmatrix}$$

其中 n 是输出 $\mathbf{y}(k)$ 的维数，而且每个 $\boldsymbol{\theta}_i(k)$ 的维数皆不超过 $\boldsymbol{\theta}(k)$ 的维数.

例如：线性系统的状态空间表示形式

$$\mathbf{x}(k+1) = A(k)\mathbf{x}(k) + B(k)\mathbf{u}(k) + \mathbf{v}(k) \tag{4.20}$$

其中 $\mathbf{x}(k)$ 是 n 维状态向量，$\mathbf{u}(k)$ 是 m 维控制向量，$\mathbf{v}(k)$ 是 n 维随机噪声，而

$$A(k) = \begin{pmatrix} a_{11}(k) & a_{12}(k) & \cdots & a_{1n}(k) \\ a_{21}(k) & a_{22}(k) & \cdots & a_{2n}(k) \\ \vdots & \vdots & & \vdots \\ a_{n1}(k) & a_{n2}(k) & \cdots & a_{nn}(k) \end{pmatrix}$$

$$B(k) = \begin{pmatrix} b_{11}(k) & b_{12}(k) & \cdots & b_{1m}(k) \\ b_{21}(k) & b_{22}(k) & \cdots & b_{2m}(k) \\ \vdots & \vdots & & \vdots \\ b_{n1}(k) & b_{n2}(k) & \cdots & b_{nm}(k) \end{pmatrix}$$

只要置

$$\boldsymbol{\theta}_i(k)^\tau = (a_{i1}(k), a_{i2}(k), \cdots, a_{in}(k), b_{i1}(k), \cdots, b_{im}(k))$$

$$\boldsymbol{\varphi}(k)^\tau = [\mathbf{x}(k)^\tau \vdots \mathbf{u}(k)^\tau]$$

$$\mathbf{y}_i(k) = x_i(k+1) \quad i = 1, 2, \cdots, n$$

则(4.20)式就可以写成

$$y_i(k) = \boldsymbol{\varphi}(k)^\tau \boldsymbol{\theta}_i(k) + \mathbf{v}_i(k) \qquad i = 1, 2, \cdots, n \tag{4.21}$$

这就是说，系统的模型(4.20)具有可分离参数的形式.

同样可以说明多输出线性系统的 CAR 模型

$$\mathbf{y}(k) + A_1(k)\mathbf{y}(k-1) + \cdots + A_p(k)\mathbf{y}(k-p)$$
$$= B_0(k)\mathbf{u}(k) + \cdots + B_q(k)\mathbf{u}(k-q) + \mathbf{e}(k) \tag{4.22}$$

也具有可分离参数的形式.

定义 设 S_1 和 S_2 是两个动态系统，如果在过去的输入完全相同的条件下，现在的同一个输入在两个系统中对应着相同的输出，我们就说这两个系统是输入-输出等效的，或简称等效的.

设 $S_1, S_2, \cdots, S_n$ 是 n 个单输出系统，如果一个 n 输出系统 S 的输出、输入和噪声分别由 $S_1, S_2, \cdots, S_n$ 的输出、输入和噪声所形成的向量而构成，则说系统 S 是由 $S_1, S_2, \cdots, S_n$ 并联而成

的，记作

$$S = S_1 \times S_2 \times S_3 \times \cdots \times S_n$$

设多输出系统的模型是

$$\mathbf{y}(k) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + \mathbf{v}(k)$$

其中

$$\mathbf{y}(k)^\tau = (y_1(k), y_2(k), \cdots, y_n(k))$$

$$\mathbf{v}(k)^\tau = (v_1(k), v_2(k), \cdots, v_n(k))$$

$$f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] = \begin{pmatrix} f_1[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^1(k), k] \\ f_2[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^2(k), k] \\ \vdots \\ f_n[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^n(k), k] \end{pmatrix}$$

则如下形式的系统称为广义单输出系统

$$y_i(k) = f_i[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^i(k), k] + v_i(k) \quad i = 1, 2, \cdots, n$$

$\theta^i(k)$ 的维数不超过 $\theta(k)$ 的维数.

定义 如果一个 n 输出系统能够与 n 个广义单输出系统的并联系统等效，就说这个系统是输出可分离的.

显然，每个参数可分离系统必然是输出可分离的. 所以为了寻求输出可分离的条件，只要得到参数可分离的条件就够了.

定理 4.4 设多输出系统的模型是

$$\mathbf{y}(k) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + \mathbf{v}(k)$$

并且相对于参量的估值 $\hat{\theta}(k)$，具有一致小于 ε 的后验残差. 其中

$$\mathbf{y}(k)^\tau = (y_1(k), y_2(k), \cdots, y_n(k))$$

$$\mathbf{v}(k)^\tau = (v_1(k), v_2(k), v_3(k), \cdots, v_n(k))$$

$$f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] = \begin{pmatrix} f_1[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] \\ f_2[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] \\ \vdots \\ f_n[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] \end{pmatrix}$$

如果有常数 $c > 0$，使对一切 k 皆有

$$|y_i(k)| < c \quad \text{a.s} \qquad i = 1, 2, \cdots, n$$

对一切 k、一切 $\mathbf{y}(k)$ 和 $\mathbf{u}(k)$、一切 $\beta \in R^m$(m 是 $\theta(k)$ 的维

数),皆有

$$|f_i[\mathbf{Y}_{k-1}, \mathbf{U}_k, \beta, k]| < c \quad \text{a.s} \qquad i = 1, 2, \cdots, n$$

考虑如下的 n 个广义单输出系统

$$y_i(k) = f_i[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^i(k), k] + v_i(k) \qquad i = 1, 2, \cdots, n$$

设 $\theta^i(k)$ 的估值 $\hat{\theta}^i(k)$ 是由 § 4.2 节的估值公式 (4.12) 所确定的,并对 $\delta > 0$ 一致地有

$$\lim_{k\to\infty} \frac{\nabla_{\hat{\theta}^i(k-1)} f_i[k, \overline{\theta^i(k)}]^\tau \nabla_{\hat{\theta}^i(k-1)} f_i[k, \hat{\theta}^i(k-1)]}{\|\nabla_{\hat{\theta}^i(k-1)} f[k, \hat{\theta}^i(k-1)]\|^2} = \mu_i > 0$$

其中 $\overline{\theta^i(k)}$ 满足

$$f_i[k, \hat{\theta}^i(k)] = f_i[k, \hat{\theta}^i(k-1)] + \nabla_{\hat{\theta}^i(k-1)} f_i[k, \overline{\theta^i(k-1)}]^\tau \tilde{\theta}^i(k)$$

$$\tilde{\theta}^i(k) = \hat{\theta}^i(k) - \hat{\theta}^i(k-1)$$

则必有 $\delta_i > 0$ 和 $N > 0$, 使得与之相应的估值序列 $\{\hat{\theta}^i(k)\}$,当 $K \geqslant N$ 时恒有

$$|y_i(k) - f_i[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}^i(k), k]| < \frac{\varepsilon}{\sqrt{n}} \qquad \text{a.s}$$

证明 由于对每个系统

$$y_i(k) = f_i[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^i(k), k] + v_i(k)$$

定理 4.1 的条件皆满足. 所以只需应用这个定理,即可得出

$$|y_i(k) - f_i[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}^i(k), k]| < \frac{\varepsilon}{\sqrt{n}} \qquad \text{a.s} \qquad \blacksquare$$

显然,系统的模型

$$\begin{pmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_n(k) \end{pmatrix} = \begin{pmatrix} f_1[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^1(k), k] \\ f_2[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^2(k), k] \\ \vdots \\ f_n[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta^n(k), k] \end{pmatrix} + \begin{pmatrix} v_1(k) \\ v_2(k) \\ \vdots \\ v_n(k) \end{pmatrix}$$

是参数可分离的,从而是输出可分离的. 上述定理说明,在保证后验残差充分一致小的前提下,只要相应的条件满足,就可以找到一个与原系统"等效"(即有同样一致小的后验残差)的输出可分离

系统。

§4.5 参数估计算法的改进

在这一节里，我们以线性情形为例来说明前面所述参数估计算法的缺欠及改进的一种途径．考虑系统

$$y_k = \boldsymbol{\phi}(k)^\tau \theta(k) + e_k$$

y_k 是一维的输出，对于参量 $\theta(k)$，我们有如下的估值算法

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{\delta}{\|\boldsymbol{\phi}(k)\|^2} \boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\} \tag{4.23}$$

当取 $\delta = 1$ 时，由上述算法所得到的估值 $\hat{\boldsymbol{\theta}}(k)$ 满足

$$y_k = \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k) \tag{4.24}$$

但是不难看出，对于初值 $\hat{\boldsymbol{\theta}}(0)$ 的不同取法，虽然观测数据相同，但所得到的估值序列 $\{\hat{\boldsymbol{\theta}}(k)\}$ 通常是不同的．然而它们都将使(4.24)式成立．这说明，算法(4.23)是受它的初值影响的，从建模的角度来看，这当然是一种缺欠．但是由于所得的估值 $\hat{\boldsymbol{\theta}}(k)$ 总能使(4.24)式成立，所以这种缺欠对预报问题来说，不会发生本质的影响．即使如此，我们仍希望能够克服这一缺欠．从而引入如下指标函数

$$J = \sum_{k=1}^{N} \|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k-1)\|^2 \tag{4.25}$$

其中 N 是观测数据的组数．由于

$$\hat{\boldsymbol{\theta}}(1) = \hat{\boldsymbol{\theta}}(0) + \frac{1}{\|\boldsymbol{\phi}(1)\|^2} \boldsymbol{\phi}(1)\{y_1 - \boldsymbol{\phi}(1)^\tau \hat{\boldsymbol{\theta}}(0)\} = F_1[\hat{\boldsymbol{\theta}}(0)]$$

$$\begin{aligned}
\hat{\boldsymbol{\theta}}(2) &= \hat{\boldsymbol{\theta}}(1) + \frac{1}{\|\boldsymbol{\phi}(2)\|^2} \boldsymbol{\phi}(2)\{y_2 - \boldsymbol{\phi}(2)^\tau \hat{\boldsymbol{\theta}}(1)\} \\
&= F_1[\hat{\boldsymbol{\theta}}(0)] + \frac{1}{\|\boldsymbol{\phi}(2)\|^2} \boldsymbol{\phi}(2)\{y_2 - \boldsymbol{\phi}(2)^\tau F_1[\hat{\boldsymbol{\theta}}(0)]\} \\
&= F_2[\hat{\boldsymbol{\theta}}(0)]
\end{aligned}$$

$\cdots$

$$\hat{\boldsymbol{\theta}}(N) = \hat{\boldsymbol{\theta}}(N-1) + \frac{1}{\|\boldsymbol{\phi}(N)\|^2}\boldsymbol{\phi}(N)\{y_N - \boldsymbol{\phi}(N)^\tau\hat{\boldsymbol{\theta}}(N-1)\}$$

$$= F_{N-1}[\hat{\boldsymbol{\theta}}(0)] + \frac{1}{\|\boldsymbol{\phi}(N)\|^2}\boldsymbol{\phi}(N)\{y_N - \boldsymbol{\phi}(N)^\tau F_{N-1}[\hat{\boldsymbol{\theta}}(0)]\}$$

$$= F_N[\hat{\boldsymbol{\theta}}(0)]$$

其中 $F_1(\cdot), F_2(\cdot), \cdots, F_N(\cdot)$ 皆为适当的函数. 由此可见，在观测数据已给定的情况下，指标函数 J 仅仅依赖于 $\hat{\boldsymbol{\theta}}(0)$，即

$$J = J[\hat{\boldsymbol{\theta}}(0)]$$

于是我们引入如下最佳初值的概念，如果 $\hat{\boldsymbol{\theta}}^*(0)$ 满足

$$J[\hat{\boldsymbol{\theta}}^*(0)] = \min_{\boldsymbol{\theta}(0)} J[\hat{\boldsymbol{\theta}}(0)]$$

我们就说 $\hat{\boldsymbol{\theta}}^*(0)$ 是最佳初值.

寻求最佳初值的一个可行的方法是最优化技术中的搜索法. 每进行一次搜索，都要从头至尾地计算一次估值序列 $\{\hat{\boldsymbol{\theta}}(k)\}$. 所以这种方法只能离线地进行.

另一方面，由于我们有

$$\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k-1) = \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1)\}$$

所以指标函数 J 又可表示为

$$J = \sum_{k=1}^{N} \frac{1}{\|\boldsymbol{\phi}(k)\|^2}[y_k - \boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1)]^2$$

特别地，如果参数 $\boldsymbol{\theta}(k) = \boldsymbol{\theta}$ 不是时变的，则我们可以近似地有

$$J = \sum_{k=1}^{N} \frac{1}{\|\boldsymbol{\phi}(k)\|^2}[y_k - \boldsymbol{\phi}(k)^\tau\boldsymbol{\theta}]^2 = J(\boldsymbol{\theta})$$

由此可见，选取 $\hat{\boldsymbol{\theta}}(0)$ 使 J 达到最小，实质上是寻求 $\boldsymbol{\theta}$ 的最小二乘估值. 这种方法当然可以在线地进行. 如果 e_k 是零均值的白噪声. 我们可以用下述递推算法来代替关于最佳估值的选取

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + M(k)\{y_k - \boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1)\}$$

$$M(k) = \frac{P(k-1)\boldsymbol{\phi}(k)}{\lambda + \boldsymbol{\phi}(k)^\tau P(k-1)\boldsymbol{\phi}(k)}$$

$$P(k)=\frac{1}{\lambda}[I-M(k)\phi(k)^{\tau}]P(k-1)$$

其中 λ 是遗忘因子.

其实我们用上述这一组递推算法所得到的是系统参数 $\theta(k)$ 的一系列估值. 根据最小二乘法的基本性质知,这组估值 $\{\hat{\theta}(k)\}$ 所揭示的是 $\theta(k)$ 的大范围变化的平均趋势. 特别这组算法所得到的估值,在时刻足够大时,与初值的选择无关. 它的缺欠是所得到的估值 $\{\hat{\theta}(k)\}$ 不能反映出 $\theta(k)$ 的瞬时变化特征. 然而这正是算法

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{1}{\|\phi(k)\|^2}\phi(k)\{y(k)-\phi(k)^{\tau}\hat{\theta}(k-1)\}$$

所能够做到的. 这就启发我们把这两组算法联合起来, 从而得出如下一组新算法

$$\hat{\theta}(k)=\hat{\theta}_1(k)+\frac{1}{\|\phi(k)\|^2}\phi(k)\{y_k-\phi(k)^{\tau}\hat{\theta}_1(k)\} \quad (4.26)$$

$$\hat{\theta}_1(k)=\hat{\theta}_1(k-1)+M(k)\{y_k-\phi(k)^{\tau}\hat{\theta}_1(k-1)\} \quad (4.27)$$

$$M(k)=\frac{P(k-1)\phi(k)}{\lambda+\phi(k)^{\tau}P(k-1)\phi(k)} \quad (4.28)$$

$$P(k)=\frac{1}{\lambda}[I-M(k)\phi(k)^{\tau}]P(k-1) \quad (4.29)$$

λ 是遗忘因子.

由于有了算法 (4.27)～(4.29), 所以这组算法已不受初值选取的影响了, 并且反映了参数的大范围变化趋势. 由于有算法 (4.26),所以这组算法也反映了 $\theta(k)$ 的瞬时变化特征. 特别地,我们仍有

$$\varepsilon(k,\hat{\theta}(k))=y_k-\phi(k)^{\tau}\hat{\theta}(k)=0$$

这只需注意到

$$\begin{aligned}&y_k-\phi(k)^{\tau}\hat{\theta}(k)\\=&y_k-\phi(k)^{\tau}\hat{\theta}_1(k)-\phi(k)^{\tau}\frac{1}{\|\phi(k)\|^2}\phi(k)\{y_k-\phi(k)^{\tau}\hat{\theta}(k)\}\end{aligned}$$

$$= y_k - \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\theta}}_1(k) - \{y_k - \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\theta}}_1(k)\} = 0$$

所以这仍然是一组适用于预报问题的参数辨识算法.

对于非线性情形,就没有类似于(4.26)～(4.29)的递推公式.然而前面所提出的离线方法仍然是适用的.

§4.6 参数估计算法的分析

在前面一节，我们已引出了指标函数

$$J_N = \sum_{k=1}^{N} \|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k-1)\|^2 \tag{4.30}$$

其中N是观测数据的组数．借助于这个指标函数，我们引出了参数估值算法的最佳初值概念．以下我们将看到，这个指标函数还有着另外的重要意义．

这里仅考虑对于系统

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k) + e(k)$$

的参数估计问题,可以证明:

引理 4.5 在 y_k, $\boldsymbol{\phi}(k)$ 和 $\hat{\boldsymbol{\theta}}(k-1)$ 给定的条件下，由下述算法确定的 $\hat{\boldsymbol{\theta}}(k)$

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k-1)\} \tag{4.31}$$

满足约束条件

$$y_k = \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k)$$

并使指标函数

$$J = \|\hat{\boldsymbol{\theta}}(k) - \boldsymbol{\theta}(k-1)\|^2$$

达到最小值.

证明 用拉格朗日乘子法．考虑函数

$$J^* = \|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k-1)\|^2 + \lambda(y_k - \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k))$$

于是有

$$\frac{\partial J^*}{\partial \hat{\boldsymbol{\theta}}(k)} = 2(\boldsymbol{\theta}(k) - \boldsymbol{\theta}(k-1)) - \lambda \boldsymbol{\phi}(k) = 0 \tag{4.32}$$

$$\frac{\partial J^*}{\partial \lambda} = y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k) = 0 \tag{4.33}$$

由(4.32)式得

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{\lambda}{2}\boldsymbol{\phi}(k) \tag{4.34}$$

同时有

$$2\boldsymbol{\phi}(k)^\tau(\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k-1)) = \lambda\|\boldsymbol{\phi}(k)\|^2$$

注意到

$$\boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k) = y_k$$

所以有

$$\frac{\lambda}{2} = \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\}$$

由(4.34)式立即得出

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\}$$

这就证明了我们的引理. ■

推论 4.6 在观测数据及初值 $\hat{\boldsymbol{\theta}}(0)$ 皆已给定的条件下，递推算法(4.31)所确定的估值序列 $\hat{\boldsymbol{\theta}}(1), \hat{\boldsymbol{\theta}}(2), \cdots, \hat{\boldsymbol{\theta}}(N)$ 必满足约束条件

$$y_k = \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k) \qquad k = 1, 2, \cdots, N$$

并使指标函数

$$J_N = \sum_{k=1}^{N} \|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k-1)\|^2$$

达到最小.

由递推算法

$$\hat{\boldsymbol{\theta}}(k) = \boldsymbol{\theta}(k-1) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\}$$

立即得出

$$\|\hat{\boldsymbol{\theta}}(k)-\hat{\boldsymbol{\theta}}(k-1)\|^2=\frac{1}{\|\boldsymbol{\phi}(k)\|^2}(y_k-\boldsymbol{\phi}(k)^\tau\boldsymbol{\theta}(k-1))^2$$

所以指标函数(4.30)又可以写成

$$J_N=\sum_{k=1}^{N}\frac{1}{\|\boldsymbol{\phi}(k)\|^2}(y_k-\boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1))^2$$

于是我们得到了如下的定理

定理 4.7 由递推算法(4.31)所确定的估值序列，在观测数据与初始估值 $\hat{\boldsymbol{\theta}}(0)$ 给定的条件下，必满足约束条件

$$y_k=\boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k)\qquad k=1,2,\cdots,N$$

并使指标函数

$$J_N=\sum_{k=1}^{N}\frac{1}{\|\boldsymbol{\phi}(k)\|^2}(y_k-\boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1))^2$$

达到最小。

在这个定理的启示下，我们可以引入关于系统

$$y_k=\boldsymbol{\phi}(k)^\tau\boldsymbol{\theta}(k)+e(k)$$

的时变参数 $\boldsymbol{\theta}(k)$ 的估计准则如下：使

$$J_N^c=\sum_{k=1}^{N}W(k)(y_k-\boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1))^2$$

达到最小。其中 $W(k)$ 是正的加权因子。不妨称 J_N^c 为加权残差平方和准则。显然，它可以看成是加权最小二乘准则向时变参数情形的推广，而我们所强调的递推算法(4.31)在约束条件

$$y_k=\boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k)\qquad k=1,2,\cdots,N$$

下，恰好满足准则

$$J_N^c=\min$$

此时取加权因子 $W(k)=\dfrac{1}{\|\boldsymbol{\phi}(k)\|^2}$。

上述的分析说明，递推算法(4.31)是可能具有较好的性质的。

下面我们转向讨论随机动态系统的参量随机化问题，这是多层建模方法的基础。

定理 4.8 对于含有未知随机部分的系统

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\eta}(k) + v(k) \tag{4.35}$$

其中 $\boldsymbol{\phi}(k)$ 是由某些输入和输出构成的向量 $\boldsymbol{\eta}(k)$ 是时变参数，$v(k)$ 是未知随机部分.则必存在随机时变参数 $\boldsymbol{\theta}(k)$，使得系统

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k)$$

与系统(4.35)是输入-输出等效的.

证明 只需置

$$\boldsymbol{\theta}(k) = \boldsymbol{\eta}(k) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)v(k)$$

就行了. 事实上我们有

$$\begin{aligned}\boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k) &= \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\eta}(k) + \boldsymbol{\phi}(k)^{\tau}\frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)v(k) \\ &= \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\eta}(k) + v(k)\end{aligned}$$

■

在进行多层建模时，如果第一层模型结构不清楚，根据这个定理可以把模型取成下述形式

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k)$$

其中 $\boldsymbol{\phi}(k)$ 是某些适当的输入和输出构成的向量，$\boldsymbol{\theta}(k)$ 是随机时变参数.

根据上述定理，我们总可以用具有随机参数 $\boldsymbol{\theta}(k)$ 的系统

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k)$$

在等效的意义下来代替具有附加噪声 $v(k)$ 的系统

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\eta}(k) + v(k)$$

我们把这种代替手续称为"关于附加噪声的参数随机化补偿".

可以证明，在适当的条件下，具有附加噪声的非线性系统

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\eta}(k), k] + v(k)$$

也可以在等效的意义下，用具有随机时变参数 $\boldsymbol{\theta}(k)$ 的系统

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\theta}(k), k]$$

来代替. 事实上，我们有下述定理:

定理 4.9 对于含有非明确随机部分的系统

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \boldsymbol{\eta}(k), k] + v(k) \tag{4.36}$$

如果它满足下述条件：有常数 $c>0$，使得对一切 k 皆有

$$|y(k)|<c \quad \text{a.s}$$

对一切 k，一切 $y(k)$ 和 $u(k)$，一切 $\beta \in R^m$ 皆有

$$|f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \beta, k]|<c \quad \text{a.s}$$

$\{\eta(k)\}$ 是由下述递推算法所确定的估值序列

$$\begin{aligned}\hat{\eta}(k)=\hat{\eta}(k-1)+&\frac{\delta}{\|V_{\hat{\eta}(k-1)}f[k, \hat{\eta}(k-1)]\|^2}\\&\times V_{\hat{\eta}(k-1)}f[k, \hat{\eta}(k-1)]\\&\cdot\{y_k-f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \hat{\eta}(k-1), k]\}\end{aligned} \tag{4.37}$$

其中 δ 是适当的常数，而

$$V_{\hat{\eta}(k-1)}f[k, \hat{\eta}(k-1)]=\frac{\partial}{\partial\beta}f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \beta, k]|_{\beta=\hat{\eta}(k-1)}$$

如果对 $\delta>0$ 一致地有

$$\hat{\eta}\|(k)\|<c \qquad \forall k \text{ a.s}$$

以及

$$0<\mu_1<\frac{V_{\hat{\eta}(k-1)}f[k, \overline{\eta(k)}]^\tau V_{\hat{\eta}(k-1)}f[k, \hat{\eta}(k-1)]}{\|V_{\hat{\eta}(k-1)}f[k, \hat{\eta}(k-1)]\|^2}<\mu_2$$

$$\forall k \text{ a.s} \tag{4.38}$$

其中 μ_1，μ_2 是常数，$\overline{\eta(k)}$ 由下式决定

$$f[k, \hat{\eta}(k)]=f[k, \hat{\eta}(k-1)]+V_{\hat{\eta}(k-1)}f[k, \overline{\eta(k)}]^\tau \tilde{\eta}(k)$$

$$\tilde{\eta}(k)=\hat{\eta}(k)-\hat{\eta}(k-1)$$

$$f[k, \eta]=f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \eta, k]$$

则必存在时变参数序列 $\{\theta(k)\}$，使得系统

$$y_k=f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \theta(k), k] \tag{4.39}$$

与系统(4.36)等价。

证明 对 $f[k, \hat{\eta}(k)]=f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \hat{\eta}(k), k]$ 在 $\hat{\eta}(k-1)$ 处应用微分中值公式，则对每个样本点 ω，有

$$f[k, \hat{\eta}(k)]=f[k, \hat{\eta}(k-1)]+V_{\hat{\eta}(k-1)}f[k, \overline{\eta(k)}]^\tau \tilde{\eta}(k)$$

但是

$$\hat{\eta}(k) = \hat{\eta}(k) - \hat{\eta}(k-1) = \hat{\eta}(k-1)$$

$$+ \frac{\delta}{\|V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]\|^2} V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]$$

$$\cdot \{y(k) - f[k, \hat{\eta}(k-1)]\} - \hat{\eta}(k-1)$$

所以对每个样本点 ω，有

$$f[k, \hat{\eta}(k)] = f[k, \hat{\eta}(k-1)]$$

$$+ \delta \frac{V_{\hat{\eta}(k-1)} f[k, \overline{\eta(k)}]^{\tau} V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]}{\|V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]\|^2}$$

$$\times \{y(k) - f[k, \hat{\eta}(k-1)]\}$$

于是

$$y_k - f[k, \hat{\eta}(k)]$$

$$= \left(1 - \delta \cdot \frac{V_{\hat{\eta}(k-1)} f[k, \overline{\eta(k)}]^{\tau} V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]}{\|V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]\|^2}\right)$$

$$\cdot \{y(k) - f[k, \hat{\eta}(k-1)]\}$$

由于有条件(4.39)式，所以必有 $\delta > 0$，使得对于任何的 $\varepsilon > 0$，有

$$0 < 1 - \delta \cdot \frac{V_{\hat{\eta}(k-1)} f[k, \overline{\eta(k)}]^{\tau} f V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]}{\|V_{\hat{\eta}(k-1)} f[k, \hat{\eta}(k-1)]\|^2} < \frac{\varepsilon}{2c}$$

$$\forall k \quad \text{a.s}$$

所以对这样的 δ 我们有

$$|y_k - f[k, \hat{\eta}(k)]| < \varepsilon \qquad \forall k \quad \text{a.s}$$

取 $\varepsilon_n \downarrow 0 (n \to \infty)$ 于是必有 δ_n，使得对一切 k 皆有

$$|y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \hat{\eta}_n(k), k]| < \varepsilon_n \ \text{a.s}$$

其中 $\eta_n(k)$ 是由算法 (4.38) 取 $\delta = \delta_n$ 时所得到的参数估值序列。由于 $\{\hat{\eta}_n(k)\}$ 以概率 1 一致有界，从而可选出它的一个子序列。不妨仍记为 $\{\hat{\eta}_n(k)\}$ 它满足

$$\lim_{n \to \infty} \hat{\eta}_n(k) = \theta(k) \qquad \text{a.s}$$

对于这样的 $\theta(k)$，显然有

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_{k-1}, \theta(k), k] \qquad \text{a.s} \qquad \blacksquare$$

§4.7 预报模型参数的完全辨识

以上我们所提出的参数跟踪(估值)算法，对于时变参数预报模型而言，仅是参数辨识的一个重要组成部分. 它只解决了依据已有的资料，对系统的"过去"和"现在"的参数值进行估计的问题. 预报模型参数辨识，特别是对时变参数的辨识，要求我们掌握参数的"全部"值. 所以我们认为，参数辨识手续应由两部分组成，即：1）依据观测数据对参数的值进行跟踪估计；2）对参数的将来值进行预测.

以下分两种情形进行讨论：

1. 非时变参数的情形

我们可以把非时变参数看成时变参数的特例. 但是非时变参数，并不存在预测问题，所以应该把它与时变参数区别开来.

设 $\hat{\boldsymbol{\theta}}(1), \hat{\boldsymbol{\theta}}(2), \cdots, \hat{\boldsymbol{\theta}}(N)$ 是依据N组观测数据所得出的关于参量的跟踪估值.

如果系统是无噪声的，并且观测数据也是准确的，则只有当

$$\hat{\boldsymbol{\theta}}(1) = \hat{\boldsymbol{\theta}}(2) = \cdots = \hat{\boldsymbol{\theta}}(N)$$

时，系统的参量才是非时变的. 此时这些估值的共同值，就是参量的最佳估值.

如果系统中存在有噪声，关于参数的非时变性检验，可以用统计假设检验方法来完成. 这里，我们仅能考虑系统噪声为零均值的正态白噪声的情形.

作统计假设 H_0: 系统的参数是非时变的.

在这一假设下，当 k 足够大时，$\hat{\boldsymbol{\theta}}(k-1)$ 可以作为非时变参量的较好估值. 从而残差

$$\varepsilon(k, \hat{\boldsymbol{\theta}}(k-1)) = y(k) - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k-1), k]$$

可以看成是噪声 $v(k)$ 的估值. $\{\varepsilon(k, \hat{\boldsymbol{\theta}}(k-1))\}$ 可以看成是从已知正态母体中的抽样值.

置

$$\hat{\boldsymbol{\theta}} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}(k)$$

以及

$$\varepsilon(k,\hat{\boldsymbol{\theta}}) = y(k) - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}, k]$$

在参数为非时变的假设下，$\{\varepsilon(k,\hat{\boldsymbol{\theta}})\}$ 与 $\{\varepsilon(k,\hat{\boldsymbol{\theta}}(k-1)\}$ 应来自同一正态母体. 于是我们可以采用熟知的统计假设检验方法，例如 χ^2-检验法进行检验. 如果接受假设 H_0，则 $\hat{\boldsymbol{\theta}}$ 就是参数的最佳估值.

2. 时变参数的情形

如果经过检验已认定参数是时变的，为进行关于参数的完全辨识，我们必须依据估值序列

$$\hat{\boldsymbol{\theta}}(1), \hat{\boldsymbol{\theta}}(2), \cdots, \hat{\boldsymbol{\theta}}(N)$$

来建立时变参数 $\boldsymbol{\theta}(k)$ 的预报估值算法，并依据这一算法寻求它的一系列预报估值，这是较复杂的问题. 对此我们可以采用下面一章中所提出的多层时间序列分析和多层递阶预报方法.

§ 4.8 典型形式 $y_k = \boldsymbol{\varphi}(k)^{\tau} A\mathbf{f}(k) + \boldsymbol{w}(k)$

从本节开始，我们将在较一般的意义下讨论动态系统时变参数的辨识问题. 关于这个问题的重要意义是众所周知的. 目前在这个方向上的研究成果，应该说还不很多. 至今可以认为已经得出了可用的关于时变参数的辨识方法，大致包括两种：一种是对于参数慢时变的情形，可以采用具有遗忘因子的递推算法或调整增益的算法来估计时变参数；另一种情形是要求已知参数变化所满足的某种模型.

然而，实际工程中所提出的时变参数系统的辨识问题，往往并不能事先知道时变参数所满足的模型，有时也无法知道它到底是否为慢时变的. 实际问题往往要求我们仅依据一系列量测数据，

来估计时变参数的未知值．但由于参数的时变特性，不同的时刻所得到的量测值只能为相应时刻的参数值的估计提供信息．即时变参数的每个瞬时时刻的估值，仅能依据对应时刻的量测数据．这当然不能满足关于非时变参数辨识时，对量测数据的“充分丰富”的要求．这是用某些已有的方法来估计时变参数的值时所遇到的本质性的困难．所以在某种特殊的场合，例如象我们在本章前几节所做的那样不得不采取一种“权宜”的算法．很明显，这种“权宜”的算法，仅能满足某种特殊性问题的要求，而不具有一般的适用性．无论对理论的要求，还是就实际应用的目的而论，寻求关于动态系统时变参数的适用于较一般情形的估值算法，当然是有着重要意义的．

这里仅考虑如下形式的系统

$$y_k = \boldsymbol{\varphi}(k)^\tau \boldsymbol{\theta}(k) + e_k \tag{4.40}$$

此处 y_k 是一维的输出，$\boldsymbol{\varphi}(k)$ 是由输入、输出数据构成的可量测的向量，$\boldsymbol{\theta}(k)$ 是时变参数，e_k 是随机噪声．我们的目的是寻求一种具有一般性的关于 $\boldsymbol{\theta}(k)$ 的估值的途径．为此，在以下引入了一种具有非时变参数的特殊形式的模型

$$y_k = \boldsymbol{\varphi}(k)^\tau A\mathbf{f}(k) + w(k) \tag{A}$$

并将其称为典型形式(A)．

其中 y_k 是一维的输出，$\boldsymbol{\varphi}(k)$ 是由输入和输出的某些数据构成的可量测的向量，不妨设其维数为 m，$w(k)$ 是随机噪声，A 是 $m \times n$ 的定常矩阵

$$A = \begin{pmatrix} a_{11} & a_{12} \cdots a_{1n} \\ a_{21} & a_{2l} \cdots a_{2n} \\ \vdots & \vdots \quad\quad \vdots \\ a_{m1} & a_{m2} \cdots a_{mn} \end{pmatrix}$$

$a_{ij}(i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, n)$ 是未知参数．$\mathbf{f}(k)$ 是由已知的函数列 $\{f_i(k)\}$ 中的元素依序构成的 n 维向量

$$\mathbf{f}(k) = [f_1(k), f_2(k), \cdots, f_n(k)]^\tau$$

为方便计，不妨设

$$\varphi(k) = [x_1(k), x_2(k), \cdots, x_m(k)]^\tau$$

于是典型形式(A)能够改写成

$$y_k = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ji} x_j(k) f_i(k) + w(k) \tag{4.41}$$

如果置

$$\boldsymbol{\phi}(k) = [x_1(k)f_1(k), x_1(k)f_2(k), \cdots, x_1 f_n(k), \\ x_2(k)f_1(k), \cdots, x_m(k)f_1(k), x_m(k)f_2(k), \cdots, \\ x_m(k)f_n(k)]^\tau$$

$$\boldsymbol{\eta} = [a_{11}, a_{12}, \cdots, a_{1n}, a_{21}, a_{22}, \cdots, a_{2n}, \cdots, \\ a_{m1}, a_{m2}, \cdots, a_{mn}]^\tau$$

则(4.41)式可以写成

$$y_k = \boldsymbol{\phi}(k)^\tau \boldsymbol{\eta} + w(k) \tag{4.42}$$

其中 $\boldsymbol{\eta}$ 为定常的未知参数向量. 于是依据噪声 $w(k)$ 的不同的统计特性,可以采取不同的方法对参数 $\boldsymbol{\eta}$ 进行(递推)估计. 例如,当 $w(k)$ 为白噪声的情形,$\boldsymbol{\eta}$ 可以用通常的最小二乘法进行估计,它的递推形式是

$$\begin{aligned} \hat{\boldsymbol{\eta}}(k) &= \hat{\boldsymbol{\eta}}(k-1) + M(k)\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\eta}}(k-1)\} \\ M(k) &= \frac{P(k-1)\boldsymbol{\phi}(k)}{\alpha + \boldsymbol{\phi}(k)^\tau P(k-1)\boldsymbol{\phi}(k)} \\ P(k) &= \frac{1}{\alpha}[I - M(k)\boldsymbol{\phi}(k)^\tau]P(k-1) \end{aligned} \tag{4.43}$$

其中 $\hat{\boldsymbol{\eta}}(k)$ 表示 η 的第 k 次估值,而 α 是遗忘因子. 当然也可以采用如下的随机递推梯度算法

$$\begin{aligned} \hat{\boldsymbol{\eta}}(k) &= \hat{\boldsymbol{\eta}}(k-1) + \frac{1}{r_k}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\eta}}(k-1)\} \\ r_k &= 1 + \sum_{i=1}^{K} \|\boldsymbol{\phi}(k)\|^2 \end{aligned} \tag{4.44}$$

当 $w(k)$ 是相关噪声的情形,为了估计 $\boldsymbol{\eta}$, 根据不同的情况,可以采用广义递推最小二乘法、推广递推最小二乘法、递推极大似

然法、递推辅助变量法等等.

§ 4.9 时变参数的辨识与典型形式

这里我们的理论依据是下述的著名的维尔斯特拉斯 (Weierstrass) 定理:

定理 4.10(维尔斯特拉斯定理) 区间 $[a, b]$ 上的任一连续函数 $f(t)$ 皆可由多项式一致地逼近. 换句话说,对于任何 $\varepsilon > 0$,必有多项式

$$P(t) = b_0 + b_1 t + b_2 t^2 + \cdots + b_n t^n$$

使对 $t \in [a, b]$ 一致地有

$$|f(t) - P(t)| < \varepsilon$$

前已指出,这里仅考虑具有时变参数的动态系统

$$y_k = \varphi(k)^\tau \theta(k) + e_k$$

的参数辨识问题. 如果对 $\theta(k-1)$ 与 $\theta(k)$ 之间进行线性内插(即用"直线"段连接点 $(k-1, \theta(k-1))$ 和 $(k, \theta(k))$, 即可得出一个连续函数,所以不妨设,当 k 连续变化时,$\theta(k)$ 是 k 的一个连续函数.

此外,为了处理方便起见, 还要对自变量 k 进行适当的变换. 例如

$$t_k = \frac{1}{k+1} \qquad t_k = \frac{1}{\sqrt{k+1}}$$

等等可以把 $\{0, 1, 2, \cdots\}$ 变到区间$[0, 1]$之中. 而

$$t_k = 1 + \frac{1}{k+1} \qquad t_k = 1 + \frac{1}{\sqrt{k+1}}$$

等等. 可以把$\{0, 1, 2, \cdots\}$变到区间$[1, 2]$之中, 等等. 无论哪种情形,我们都要求

$$t_k = g(k)$$

是一个可逆的连续变换. $g^{-1}(\cdot)$ 是它的逆变换,我们引入这种变换的目的是要把较大的 k 值变换到一个较小的区间中去. 从而为

处理连续函数的多项式逼近问题带来方便。

在变换 $t_k = g(k)$ 之下，我们有

$$\theta(k) = \theta[g^{-1}(g(k))] = \theta^*[g(k)] = \theta^*(t_k)$$

这里 $\theta^*[\cdot] = \theta[g^{-1}(\cdot)]$. 显然 $\theta^*(t_k)$ 仍是 t_k 的一个连续函数. 以下为了叙述方便，置

$$\theta(k) = \begin{pmatrix} \theta_1(k) \\ \theta_2(k) \\ \vdots \\ \theta_m(k) \end{pmatrix}$$

由于每个 $\theta_i(k) = \theta_i^*(t_k)$ 都是 t_k 的连续函数. 设 t_k 的取值区间是 I. 则由维尔斯特拉斯定理对于任何的 $\varepsilon > 0$，必有 m 个多项式

$$P_i(t_k) = b_{i,0} + b_{i,1}t_k + b_{i,2}t_k^2 + \cdots + b_{i,n-1}t_k^{n-1}$$

使得对于所考虑的 k(即对于 $t_k \in I$)一致地有

$$|\theta_i^*(t_k) - P_i(t_k)| < \frac{\varepsilon}{M\sqrt{m}}$$

此处 M 为常数，它满足 $\|\varphi(k)\| < M$.

注 在应用维尔斯特拉斯定理时，我们辜且假定 t_k 是连续变化的。

如果取 $f_1(k) = 1, f_2(k) = t_k, \cdots, f_n(k) = t_k^{n-1}$，即置

$$f(k) = [1, t_k, t_k^2, \cdots, t_k^{n-1}]^\tau$$

以及

$$A = \begin{pmatrix} b_{1,0} & b_{1,1} & b_{1,2} \cdots b_{1,n-1} \\ b_{2,0} & b_{2,1} & b_{2,2} \cdots b_{2,n-1} \\ \vdots & \vdots & \vdots \quad\quad \vdots \\ b_{m,0} & b_{m,1} & b_{m,2} \cdots b_{m,n-1} \end{pmatrix}$$

则我们可以得出模型

$$y_k = \varphi(k)^\tau A\mathbf{f}(k) + w(k)$$

此处 $w(k) = e_k + \varepsilon(k)$，$\varepsilon(k)$是由 $\varphi(k)^\tau A\mathbf{f}(k)$ 代替 $\varphi(k)^\tau\theta(k)$ 所产生的误差. 这个模型已具有典型形式 (A). 这个模型与模型

$$y_k = \varphi(k)^\tau \theta(k) + e_k$$

的关系如下

$$\|\theta(k) - A\mathbf{f}(k)\| < \frac{\varepsilon}{M}$$

$$\begin{aligned} |w(k) - e_k| &= |\varphi(k)^\tau \theta(k) - \varphi(k)^\tau A\mathbf{f}(k)| \\ &= |\varphi(k)^\tau (\theta(k) - A\mathbf{f}(k))| \\ &\leqslant \|\varphi(k)\| \|\theta(k) - A\mathbf{f}(k)\| < \varepsilon \end{aligned}$$

综合上述，我们已经证明了下述的定理：

定理 4.11 对于具有时变参数的模型

$$y_k = \varphi(k)^\tau \theta(k) + e_k$$

假定有 $M > 0$，使 $\|\varphi(k)\| < M$. 则对于任何的 $\varepsilon > 0$ 必有具有典型形式(A)的模型

$$y_k = \varphi(k)^\tau A\mathbf{f}(k) + w(k)$$

使得

$$\|\theta(k) - A\mathbf{f}(k)\| < \frac{\varepsilon}{M}$$

以及

$$|\varphi(k)^\tau \theta(k) - \varphi(k)^\tau A\mathbf{f}(k)| = |w(k) - e(k)| < \varepsilon$$

从上述的定理可以看出：在近似的意义下，关于模型

$$y_k = \varphi(k)^\tau \theta(k) + w(k) \tag{4.45}$$

的时变参数 $\theta(k)$ 的辨识问题，可以转化为具有典型形式 (A) 的模型

$$y_k = \varphi(k)^\tau A\mathbf{f}(k) + e_k \tag{4.46}$$

的非时变参数矩阵 A 的辨识问题. 而且可使上述具有典型形式(A) 的模型(4.46)按任意要求的精度逼近原来的模型(4.45).

由于在模型 (4.46) 中我们已把随机噪声换成了 e_k，而不是 $w(k)$，所以它是模型(4.45)的一种近似.

注 一般的 $\varphi(k)$ 是由某些输入和输出构成的 m 维的向量，所以定理中的条件，$\|\varphi(k)\| < M$，相当于要求系统的输入、输出皆有界.

§ 4.10 典型形式的结构确定

当 $\boldsymbol{\varphi}(k)$ 的维数 m 为已知的情况下，所谓确定典型形式（A）

$$y_k = \boldsymbol{\varphi}(k)^\tau A\mathbf{f}(k) + e_k \tag{4.47}$$

的结构，事实上就是确定矩阵 A 的列数 n，也就是确定向量 $\mathbf{f}(k)$ 的维数 n.

显然，当观测数据

$$(y_1, \boldsymbol{\varphi}(1)), (y_2, \boldsymbol{\varphi}(2)), \cdots, (y_N, \boldsymbol{\varphi}(N)) \tag{4.48}$$

已知时，只要 n 确定了，$\mathbf{f}(k)$ 也就随之确定了. 所以残差平方和

$$\sum_{k=1}^{N} (y_k - \boldsymbol{\varphi}(k)^\tau A\mathbf{f}(k))^2$$

仅依赖于 A 和 n. 我们把它记为 $J(n, A)$，即

$$J(n,A) = \sum_{k=1}^{N} (y_k - \boldsymbol{\varphi}(k)^\tau A\mathbf{f}(k))^2$$

以后把与 n 对应的参数阵 A 记成 A_n，令

$$J(n) = \min_{A_n} J(n, A_n) \tag{4.49}$$

由 $J(n)$ 的定义不难得出

$$J(n) \geqslant J(n+1)$$

事实上，如果置 $\hat{A}_n$ 满足

$$J(n) = J(n, \hat{A}_n) = \min_{A_n} J(n, A_n)$$

设 $\mathbf{0}$ 是 m 维的零向量，置

$$A_{n+1}^* = [\hat{A}_n | \mathbf{0}]$$

则我们有

$$J(n+1, A_{n+1}^*) = J(n, \hat{A}_n)$$

所以

$$J(n) = J(n+1, A_{n+1}^*) \geqslant \min_{A_{n+1}} J(n+1, A_{n+1}) = J(n+1)$$

由上述分析可见，如果 $\hat{n}$ 存在，则对于充分大的 n 必有

$$J(n) = J(n+1)$$

所以最佳的 $\hat{n}$ 必满足

$$J(\hat{n}-1) > J(\hat{n}) = J(\hat{n}+1)$$

由此可得出一个对 $\hat{n}$ 进行逐步搜索的方法，函数 $J(n)$，可以作为确定参数 n 的一个准则函数．应用这一准则函数确定的最佳的 n 值 $\hat{n}$ 满足

$$J(\hat{n}) = \min_{n} J(n) = \min_{n} \min_{A_n} J(n, A_n)$$

以下我们指出，在一定的条件下，也可以采用假设检验的方法来确定 n：

如果假定 $\{e_k\}$ 是白噪声序列． 那么对于 n 的最佳估值 $\hat{n}$，残差

$$\varepsilon(k) = y_k - \boldsymbol{\varphi}(k)^{\tau} \hat{A} \mathbf{f}(k) \qquad k = 1, 2, \cdots, N$$

应该是白噪声的一系列"量测"值．其中 $\hat{A}$ 表示 A 的最佳估值．所以，我们可以借助于这些值来检验如下的统计假设：

H_0: e_k 之间是相互独立(或不相关)的．

如果用 $\hat{\rho}_k$ 表示 e_k 的相关函数的估值即

$$\hat{\rho}_k = \frac{\sum_{i=k+1}^{N} \varepsilon(i)\varepsilon(i-k)}{\sum_{i=k+1}^{N} \varepsilon(i-k)^2} \tag{4.50}$$

那么检验前述的统计假设，相当于检验

$$E\{\hat{\rho}_k\} = 0$$

因为 $\hat{\rho}_k$ 是渐近正态的(参看文献[8]或[9])．而且 $\hat{\rho}_k$ 的方差近似地有

$$\sigma(\hat{\rho}_k) \approx \frac{1}{\sqrt{N}}$$

故当置信度取成 5% 时，其置信区间近似地为

$$I_{0.05} = \left[-\frac{2}{\sqrt{N}}, \frac{2}{\sqrt{N}}\right]$$

于是得出了检验 $\hat{n}$ 的步骤如下：

取定 $\hat{n}$，计算相应的 $\varepsilon(k)$，进而求出 $\hat{\rho}_k$；

如果 $\hat{\rho}_k \in I_{0.05}$，就接受 $\{\varepsilon(k)\}$ 为白噪声这一假设，从而接受 $\hat{n}$ 为 n 的最佳估计值。

如果 $\hat{\rho}_k \notin I_{0.05}$ 就拒绝上述假设．此时就需对 $\hat{n}$ 的新值进行检验。

在 $\boldsymbol{\varphi}(k)$ 的维数 m 亦未知的情形，当观测数据（4.48）已给定时，残差平方和

$$\sum_{k=1}^{N}(y_k-\boldsymbol{\varphi}(k)^{\tau}A\mathbf{f}(k))^2$$

除依赖 n，A 外，还要依赖于 m．故它可记成 $J(m,n,A)$．令

$$J(m,n)=\min_{A} J(m,n,A)$$

此时，我们把满足下述表达式的 $\hat{m}$，$\hat{n}$ 作为 m 和 n 的最佳估值

$$J(\hat{m},\hat{n})=\min_{m,n} J(m,n)$$

显然确定这样的 $\hat{m}$，$\hat{n}$ 已经是很复杂的事了．

§4.11 算法的递推化问题

这里我们所给出的是一种在线与离线交互使用的方法。

设我们已得出了不等式

$$J(\hat{n}-1)>J(\hat{n})=J(\hat{n}+1)$$

这就是说，我们已得到了 $\hat{n}$ 和相应的参数估值 $\hat{\boldsymbol{\eta}}_{\hat{n}}$．当然同时也得到了 $\hat{\boldsymbol{\eta}}_{\hat{n}+1}$．为明确起见，我们把在（4.42）式中的与 n 相应的 $\boldsymbol{\phi}(k)$ 记成 $\boldsymbol{\phi}_n(k)$．

在线和离线的交互算法的基本步骤如下：

1．以 $\hat{\boldsymbol{\eta}}_{\hat{n}}$ 为初值，应用递推算法

$$\hat{\boldsymbol{\eta}}_{\hat{n}}(k)=\hat{\boldsymbol{\eta}}_{\hat{n}}(k+1)+M_{\hat{n}}(k)\{y_k-\boldsymbol{\phi}_{\hat{n}}(k)^{\tau}\hat{\boldsymbol{\eta}}_{\hat{n}}(k-1)\}$$

$$M_{\hat{n}}(k)=\frac{P_{\hat{n}}(k-1)\boldsymbol{\phi}_{\hat{n}}(k)}{\alpha+\boldsymbol{\phi}_{\hat{n}}(k)^{\tau}P_{\hat{n}}(k-1)\boldsymbol{\phi}_{\hat{n}}(k)}$$

$$P_{\hat{n}}(k)=\frac{1}{\alpha}[I_{\hat{n}}-M_{\hat{n}}(k)\boldsymbol{\phi}_{\hat{n}}(k)^{\tau}]P_{\hat{n}}(k-1)$$

对 $\eta_{\hat{n}}$ 进行递推估计. 其中 α 是遗忘因子，$I_{\hat{n}}$ 是阶数适当的单位矩阵.

2. 以 $\hat{\eta}_{\hat{n}+1}$ 为初值,应用递推算法

$$\hat{\eta}_{\hat{n}+1}(k)=\hat{\eta}_{\hat{n}+1}(k-1)+M_{\hat{n}+1}(k)\{y_k-\phi_{\hat{n}+1}(k)^{\tau}\hat{\eta}_{\hat{n}+1}(k-1)\}$$

$$M_{\hat{n}+1}(k)=\frac{P_{\hat{n}+1}(k-1)\phi_{\hat{n}+1}(k)}{\alpha+\phi_{\hat{n}+1}(k)^{\tau}P_{\hat{n}+1}(k-1)\phi_{\hat{n}+1}(k)}$$

$$P_{\hat{n}+1}(k)=\frac{1}{\alpha}[I_{\hat{n}+1}-M_{\hat{n}+1}(k)\phi_{\hat{n}+1}(k)^{\tau}]P_{\hat{n}}(k-1)$$

对 $\eta_{\hat{n}+1}$ 进行递推估计. 此处 $I_{\hat{n}+1}$ 是阶数适当的单位矩阵.

3. 计算 $J(\hat{n},\hat{\eta}_{\hat{n}}(k))$ 和 $J(\hat{n}+1,\hat{\eta}_{\hat{n}+1}(k))$. 此处

$$J(\hat{n},\hat{\eta}_{\hat{n}}(k))=\sum_{i=k-N}^{k}[y_i-\phi_{\hat{n}}(i)^{\tau}\hat{\eta}_{\hat{n}}(k)]^2$$

$$J(\hat{n}+1,\hat{\eta}_{\hat{n}+1}(k))=\sum_{i=k-N}^{k}[y_i-\phi_{\hat{n}+1}(i)^{\tau}\hat{\eta}_{\hat{n}+1}(k)]^2$$

其中N为正整数.

4. 作判断. 如果 $J(\hat{n},\hat{\eta}_{\hat{n}}(k))=J(\hat{n}+1,\hat{\eta}_{\hat{n}+1}(k))$，则递推算法继续进行.

如果 $J(\hat{n},\hat{\eta}_{\hat{n}}(k))>J(\hat{n}+1,\hat{\eta}_{\hat{n}+1}(k))$

则用准则函数

$$J(n,\eta_n)=\sum_{i=k-N}^{k}[y_i-\phi_n(i)^{\tau}\eta_n]^2$$

按前面所述的方法重新确定最佳的 $\hat{n}$ 和 $\hat{\eta}_{\hat{n}}$ 以及 $\hat{\eta}_{\hat{n}+1}$. 然后重复上述的递推算法.

这种算法的弱点在于要求记忆固定长度N的观测数据集,

第五章　多层递阶建模与预报

本章主要介绍多层递阶预报方法。这种方法的基本思想是把动态系统的状态预报问题分为两部分，即对系统的时变参数的预报和在此基础上对系统状态的预报. 由于这种方法充分注意到了系统的时变特性,所以预报精度有了较大的提高.

§ 5.1　时间序列的多层分析方法

设 $\{y_k\}$ 是一维时间序列，不必假定它是平稳的. 为解决这一序列的预报问题，最基本的是建立这个序列的向前一步的预报模型

$$y_k = \alpha_1(k)y_{k-1} + \alpha_2(k)y_{k-2} + \cdots + \alpha_n(k)y_{k-n} + e_k \qquad (5.1)$$

其中，$\alpha_1(k), \alpha_2(k), \cdots, \alpha_n(k)$ 为时变参数，e_k 为随机噪声，n 是模型的阶数。

置

$$\boldsymbol{\phi}(k) = [y_{k-1}, y_{k-2}, \cdots, y_{k-n}]^{\tau}$$
$$\boldsymbol{\theta}(k) = [\alpha_1(k), \alpha_2(k), \cdots, \alpha_n(k)]^{\tau}$$

则模型(5.1)可以写成

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k) + e_k \qquad (5.2)$$

我们可以用前面所指出的"参数随机化补偿"的办法,把模型 (5.2) 中的附加噪声 e_k 化到系统的时变参数上去，为此只需置

$$\boldsymbol{\beta}(k) = \boldsymbol{\theta}(k) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)e_k$$

由于

$$\boldsymbol{\phi}(k)^{\tau}\boldsymbol{\beta}(k) = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k) + \boldsymbol{\phi}(k)^{\tau}\frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)e_k$$

$$= \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\theta}(k) + e_k$$

所以有

$$y_k = \boldsymbol{\phi}(k)^{\tau}\boldsymbol{\beta}(k) \tag{5.3}$$

其中

$$\boldsymbol{\beta}(k) = [\beta_1(k), \beta_2(k), \cdots, \beta_n(k)]^{\tau}$$

是随机时变参数．显然 $\{\boldsymbol{\beta}(k)\}$ 又构成了一个 n 维的时间序列．相对于原时间序列 $\{y_k\}$ 而言，我们称 $\{\boldsymbol{\beta}(k)\}$ 为第二层时间序列．

如果对于 $\{y_k\}$ 已有足够多的观测数据

$$\{y_0, y_1, \cdots, y_N\} = \mathbf{Y}_N$$

则我们可以依据递推算法

$$\hat{\boldsymbol{\beta}}(k) = \hat{\boldsymbol{\beta}}(k-1) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\beta}}(k-1)\} \tag{5.4}$$

或者

$$\hat{\boldsymbol{\beta}}(k) = \hat{\boldsymbol{\beta}}_1(k) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\beta}}_1(k)\} \tag{5.5a}$$

$$\hat{\boldsymbol{\beta}}_1(k) = \hat{\boldsymbol{\beta}}_1(k-1) + M(k)\{y_k - \boldsymbol{\phi}(k)^{\tau}\hat{\boldsymbol{\beta}}_1(k-1)\} \tag{5.5b}$$

$$M(k) = \frac{P(k-1)\boldsymbol{\phi}(k)}{\lambda + \boldsymbol{\phi}(k)^{\tau}P(k-1)\boldsymbol{\phi}(k)} \tag{5.5c}$$

$$P(k) = \frac{1}{\lambda}[I - M(k)\boldsymbol{\phi}(k)^{\tau}]P(k-1) \tag{5.5d}$$

λ 为遗忘因子，来确定时间序列 $\{\boldsymbol{\beta}(k)\}$ 的部分估值

$$\hat{\boldsymbol{\beta}}(n), \hat{\boldsymbol{\beta}}(n+1), \cdots, \hat{\boldsymbol{\beta}}(N)$$

自然可以依据这些数据对时间序列 $\{\beta(k)\}$ 的特性进行统计分析．例如，判断它的平稳性，估计它的协方差矩阵等等．为方便计，不妨设 $\{\beta(k)\}$ 的估值序列为

$$\hat{\boldsymbol{\beta}}(0), \hat{\boldsymbol{\beta}}(1), \cdots, \hat{\boldsymbol{\beta}}(N)$$

依据这些数据，可以对关于 $\beta(k)$ 的如下的统计假设进行检验

$$H_0: \boldsymbol{\beta}(k) = \boldsymbol{\beta} + \mathbf{V}(k)$$

$\boldsymbol{\beta}$ 是非时变的定常向量，$\{\mathbf{V}(k)\}$ 是零均值的平稳随机序列. 如果接受假设 H_0，则 $\{\boldsymbol{\beta}(k)\}$ 是一个多维的平稳随机序列. 我们可以应用多维平稳序列的建模方法，建立起关于 $\{\boldsymbol{\beta}(k)\}$ 所满足的模型 (ARMA 类型的模型). 如果拒绝假设 H_0，则 $\{\boldsymbol{\beta}(k)\}$ 是一个一般的随机序列. 我们可以对 $\{\boldsymbol{\beta}(k)\}$ 建立如下的 n 个广义单输出 AR 模型

$$\beta_i(k) = A_1(k, i)^\tau \boldsymbol{\beta}(k-1) + A_2(k,i)^\tau \boldsymbol{\beta}(k-2) + \cdots + A_{n_i}(k, i)^\tau \boldsymbol{\beta}(k-n_i) \qquad (5.6)$$

$$i = 1, 2, \cdots, n.$$

其中 $A_1(k, i), A_2(k, i), \cdots, A_{n_i}(k, i)$ 皆为随机向量参数，n_i 为模型的阶数.

置

$$\boldsymbol{\gamma}_i(k) = [A_1(k, i)^\tau, A_2(k, i)^\tau, \cdots, A_{n_i}(k, i)^\tau]^\tau$$

$$i = 1, 2, \cdots, n$$

于是我们得到了 n 个多维的时间序列. 相对于原时间序列 $\{y_k\}$ 而言，称它们为第三层的时间序列.

完全类似地，我们可以用算法(5.4)或者算法(5.5)，依据数据 $\hat{\boldsymbol{\beta}}(0), \hat{\boldsymbol{\beta}}(1), \cdots, \hat{\boldsymbol{\beta}}(N)$ 来求得估值

$$\hat{\boldsymbol{\gamma}}_i(n_i), \hat{\boldsymbol{\gamma}}_i(n_i+1), \cdots, \hat{\boldsymbol{\gamma}}_i(N) \qquad i = 1, 2, \cdots, n$$

不妨设这些数据仍然从时刻 0 开始，即有

$$\hat{\boldsymbol{\gamma}}_i(0), \hat{\boldsymbol{\gamma}}_i(1), \cdots, \hat{\boldsymbol{\gamma}}_i(N) \qquad i = 1, 2, \cdots, n$$

依据这些数据，我们仍然能够对第三层时间序列 $\{\boldsymbol{\gamma}_i(k)\}$ $i = 1, 2, \cdots, n$ 的统计特性进行检验. 首先要检验的是平稳性，如果它们是平稳的，则我们可以对它们进行另外的一些分析与估计. 例如估计它们的协方差函数，估计它们的谱，分析它们的变化周期性等等. 进而我们还可以建立起这些时间序列所满足的 ARMA 模型.

如果在这些时间序列中，有的仍然是非平稳的，则我们对这种非平稳的序列，还要建立起广义单输出 AR 模型，进而得出第四层时间序列，如此等等.

一个自然的问题是，在建立各层的 AR 模型或广义单输出 AR 模型时，它们的阶数应怎样确定. 不失一般性，不妨以模型

$$y_k = \beta_1(k)^\tau y_{k-1} + \beta_2(k)^\tau y_{k-2} + \cdots + \beta_n(k)^\tau y_{k-n} \quad (5.7)$$

为例. 为确定它的适当阶数 n，令

$$J(n) = \sum_{k=n}^{N-1} \|\hat{\beta}(k+1) - \hat{\beta}(k)\|^2$$

其中 $\hat{\beta}(k)$ 由算法(5.4)或算法(5.5)来确定. 如果 n 使得

$$J(n) = \min_m J(m)$$

我们就认为它是模型(5.7)的"适当阶数".

事实上，在实际应用中，我们往往是凭经验来确定模型的"适当阶数"，它的值通常是 2～7.

由于模型的参数估值是用算法(5.4)或(5.5)来确定的，所以无论模型的阶数 n 怎样选取，都会经过时变参数值的补偿而使等式

$$y_k = \phi(k)^\tau \hat{\beta}(k)$$

成立. 由此可见，模型的阶数选取得适当与否，对于它在预报中应用的效果，不会有什么本质的影响. 而仅仅是给分析与处理的过程带来麻烦.

§5.2 多层递阶预报方法

考虑向前一步的预报误差模型

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + e_k$$

其中 y_k 是一维输出，U_k 是输入，$\theta(k)$ 是 m 维的参量，e_k 是一维噪声. 我们已经证明，对于任何的 $\varepsilon > 0$，在适当的条件下必有 $N > 0$ 和 $\delta > 0$ 存在，使得当 $k \geqslant N$ 时，由算法

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\delta}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^2}$$
$$\times \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\{y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}_{(k-1)} k]\}$$

所确定的 $\hat{\theta}(k)$ 恒满足

$$|y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k), k]| < \varepsilon$$

特别地，当

$$f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] = \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta}(k)$$

时，则由算法

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2} \boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)\}$$

所确定的 $\hat{\boldsymbol{\theta}}(k)$ 必满足

$$y_k - \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k) = 0$$

即

$$y_k = \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k)$$

上述分析结果启发我们，如果把通常的向前一步的预报公式

$$\hat{y}(k|\hat{\boldsymbol{\theta}}(k-1)) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k-1), k]$$

或

$$\hat{y}(k|\hat{\boldsymbol{\theta}}(k-1)) = \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}(k-1)$$

换成

$$\hat{y}(k|\hat{\boldsymbol{\theta}}^*(k)) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}^*(k), k]$$

或

$$\hat{y}(k|\hat{\boldsymbol{\theta}}^*(k)) = \boldsymbol{\phi}(k)^\tau \hat{\boldsymbol{\theta}}^*(k)$$

其中 $\hat{\boldsymbol{\theta}}^*(k)$ 表示依据资料 $\mathbf{Y}_{k-1}$，$\mathbf{U}_k$ 所得出的关于 $\hat{\boldsymbol{\theta}}(k)$ 的预报估值，则可使预报残差

$$\varepsilon(k, \hat{\boldsymbol{\theta}}^*(k)) = y_k - \hat{y}(k|\hat{\boldsymbol{\theta}}^*(k))$$

得到减小．这种思想不仅可以应用于向前一步的预报问题，也可以应用于向前多步的预报问题．于是我们可以提出如下的所谓“多层递阶预报方法”．这种方法的基本点是把动态系统的状态(或输出) 预报问题分为两部分：系统的时变参数预报和在此基础上对系统的状态(或输出)预报．

以一般的预报误差模型为例：

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + e_k$$

假定随机噪声 e_k 与 $f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k]$ 是独立的． 多层递阶预报方法的具体步骤是：

1. 应用适当的参数估值算法，对系统的时变参数 $\theta(k)$ 进行

估值，使得出的估值序列 $\{\hat{\boldsymbol{\theta}}(k)\}$ 满足后验残差一致小准则.

2. 依据 $\{\boldsymbol{\theta}(k)\}$ 的一系列估值: $\{\hat{\boldsymbol{\theta}}(0), \hat{\boldsymbol{\theta}}(1), \cdots, \hat{\boldsymbol{\theta}}(N)\} = \Theta_N^*$ 检验 $\{\boldsymbol{\theta}(k)\}$ 的平稳性. 如果 $\{\boldsymbol{\theta}(k)\}$ 是平稳的，则依据数据 Θ_N^* 建立 $\boldsymbol{\theta}(k)$ 所满足的多维 AR 模型或多维 ARMA 模型. 进而依据这些模型，求出 $\boldsymbol{\theta}(k)$ 的由向前一步直到向前 h 步的预报值

$$\hat{\boldsymbol{\theta}}^*(N+1), \hat{\boldsymbol{\theta}}^*(N+2), \cdots, \hat{\boldsymbol{\theta}}^*(N+h)$$

如果 $\{\boldsymbol{\theta}(k)\}$ 不具有平稳性，则可分成三种情形进行.

1) 如果能够找到一个预报误差模型

$$\boldsymbol{\theta}(k) = G[\Theta_{k-1}, \boldsymbol{\beta}, k] + \mathbf{V}(k)$$

其中参量 $\boldsymbol{\beta}$ 是非时变的，$\{\mathbf{V}(k)\}$ 是零均值的不相关随机向量序列，$\Theta_{k-1} = \{\boldsymbol{\theta}(0), \cdots, \boldsymbol{\theta}(k-1)\}$. 则我们可以依据

$$\hat{\boldsymbol{\theta}}(N+h) = G[\Theta_{N+h}, \boldsymbol{\beta}, N+h-1]$$

逐步求得 $\boldsymbol{\theta}(k)$ 的预报估值

$$\hat{\boldsymbol{\theta}}^*(N+1), \hat{\boldsymbol{\theta}}^*(N+2), \cdots, \hat{\boldsymbol{\theta}}^*(N+h)$$

2) 虽然不能够找到 $\boldsymbol{\theta}(k)$ 所满足的较好的预报公式，但能够较确切地找到时间序列 $\{\boldsymbol{\theta}(k)\}$ 的变化规律，则我们也可以依据此规律得到 $\boldsymbol{\theta}(k)$ 的预报估值

$$\hat{\boldsymbol{\theta}}^*(N+1), \hat{\boldsymbol{\theta}}^*(N+2), \cdots, \hat{\boldsymbol{\theta}}^*(N+h)$$

3) 如果没有足够把握得到 1) 或 2) 的结果，则可直接的对 $\{\boldsymbol{\theta}(k)\}$依据数据 Θ_N^* 建立广义单输出 AR 模型，继续进行关于序列 $\{\boldsymbol{\theta}(k)\}$ 的多层分析，直到某一高层的各时间序列皆为平稳序列时为止. 对于这一层的各序列建立 AR 或 ARMA 模型，用这些模型，进行相应参数的向前 h 步预报，再用这些预报值来进行关于下一层参数的向前 h 步预报. 如此下去，最后将得出 $\boldsymbol{\theta}(k)$的预报估值

$$\hat{\boldsymbol{\theta}}^*(N+1), \hat{\boldsymbol{\theta}}^*(N+2), \cdots, \hat{\boldsymbol{\theta}}^*(N+h)$$

3. 确定 Y_k 的向前 1 步至 h 步的预报估值. 此时我们有:

向前一步的预报算法

$$\hat{y}(N+1) = f[\mathbf{Y}_N, \mathbf{U}_{N+1}^*, \hat{\boldsymbol{\theta}}^*(N+1), N+1]$$

其中

$$\mathbf{Y}_N = \{y_0, y_1, \cdots, y_N\}$$

$$\mathbf{U}^*_{N+1} = \{u_0, u_1, \cdots, u_N, \hat{u}_{N+1}\}$$

$\hat{u}_{N+1}$ 表示预计的输入值. 在许多预报问题中，u_k 往往表示预报因子，所以 $\hat{u}_{N+1}$ 一般的也是用预报的手段而得到的.

向前两步的预报算法

$$\hat{y}(N+2) = f[\mathbf{Y}^*_{N+1}, \mathbf{U}^*_{N+2}, \hat{\boldsymbol{\theta}}^*(N+2), N+2]$$

其中

$$\mathbf{Y}^*_{N+1} = \{y_0, y_1, \cdots, y_N, \hat{y}(N+1)\}$$

$$\mathbf{U}^*_{N+2} = \{u_0, u_1, \cdots, u_N, \hat{u}_{N+1}, \hat{u}_{N+2}\}$$

$\hat{u}_{N+1}$, $\hat{u}_{N+2}$ 皆为预计的输入值，它们通常是用预报手段得到的.

一般地，可以得出向前 h 步的预报算法

$$\hat{y}(N+h) = f[\mathbf{Y}^*_{N+h-1}, \mathbf{U}^*_{N+h}, \hat{\boldsymbol{\theta}}^*(N+h), N+h]$$

其中

$$\mathbf{Y}^*_{N+h-1} = \{y_0, y_1, \cdots, y_N, \hat{y}_{N+1}, \cdots, \hat{y}(N+h-1)\}$$

$$\mathbf{U}^*_{N+h} = \{u_0, \cdots, u_N, \hat{u}_{N+1}, \cdots, \hat{u}_{N+h}\}$$

$\hat{u}_{N+i}$ 表示 u_N 的向前 i 步的预计(报)值.

这里，我们所提出的是向前 h 步的逐步递推预报算法.

对于特殊的情形，例如

$$\begin{aligned} f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\theta}(k), k] &= \alpha_1(k)y_{k-1} + \alpha_2(k)y_{k-2} \\ &\quad + \cdots + \alpha_n(k)y_{k-n} + \beta_0(k)u_k + \beta_1(k)u_{k-1} \\ &\quad + \cdots + \beta_m(k)u_{k-m} \end{aligned}$$

的情形，我们还可以引用奥斯特隆姆的预报方法. 关于这一点请参看 § 5.5 节.

§ 5.3 高层时间序列建模与预报的具体方法

下面介绍高层时间序列建模与预报的一些具体方法，这些方法在应用中已被证明是有效的. 由于任何高层时间序列，相对于它的前一层而言，都是第二层时间序列，所以不失一般性，我们只

考虑第二层时间序列的建模与预报就够了．设已根据观测数据，得到了第二层时间序列的估值

$$\hat{\boldsymbol{\theta}}(1), \hat{\boldsymbol{\theta}}(2), \cdots, \hat{\boldsymbol{\theta}}(N)$$

此处N是观测数据的组数．而

$$\hat{\boldsymbol{\theta}}(k)^{\tau} = (\hat{\boldsymbol{\theta}}_1(k), \hat{\boldsymbol{\theta}}_2(k), \cdots, \hat{\boldsymbol{\theta}}_m(k))$$

以下分两种情形来讨论关于 $\{\hat{\boldsymbol{\theta}}(k)\}$ 的建模和预报问题．

1. 统一处理的途径

由于 $\hat{\boldsymbol{\theta}}(k)$ 的各分量间不可避免地要存在着某些联系，所以一般情况下，为建立 $\{\boldsymbol{\theta}(k)\}$ 的预报模型，我们应该采取统一处理的途径．经常采用的是多维 AR 模型

$$\hat{\boldsymbol{\theta}}(k+1) = A(0)\hat{\boldsymbol{\theta}}(k) + A(1)\hat{\boldsymbol{\theta}}(k-1) + \cdots + A(n)\hat{\boldsymbol{\theta}}(k-n) + \mathbf{v}(k) \tag{5.8}$$

显然，可以把上述模型分解成m个广义单输出模型

$$\hat{\boldsymbol{\theta}}_i(k+1) = \mathbf{A}_i(0)\hat{\boldsymbol{\theta}}(k) + \mathbf{A}_i(1)\hat{\boldsymbol{\theta}}(k-1) + \cdots + \mathbf{A}_i(n)\hat{\boldsymbol{\theta}}(k-n) + v_i(k)$$
$$i = 1, 2, \cdots, m \tag{5.9}$$

其中 $\hat{\boldsymbol{\theta}}_i(k+1)$ 表示 $\hat{\boldsymbol{\theta}}(k+1)$ 的第 i 个分量，而 $\mathbf{A}_i(k)$ 表示矩阵 $A(k)$ 的第 i 个行向量，$v_i(k)$ 表示噪声 $\mathbf{v}(k)$ 的第 i 个分量．

我们也可以分别建立上述的m个广义单输出模型，此时它们的阶数不一定是同一个 n．

对于广义单输出模型，其未知参数向量 $\mathbf{A}_i(0)$，$\mathbf{A}_i(1)$，$\cdots$，$\mathbf{A}_i(n)$（$i = 1, 2, \cdots, m$）完全可以按我们在前章中所指出的算法来进行估计．如果经检验可以认为这些参数已是非时变的（事实上，这时序列 $\{\boldsymbol{\theta}(k)\}$ 可能已是平稳的了），则所求的向前一步的预报公式是

$$\hat{\theta}_i^*(N+1) = \hat{\boldsymbol{A}}_i(0)\hat{\boldsymbol{\theta}}(N) + \hat{\boldsymbol{A}}_i(1)\hat{\boldsymbol{\theta}}(N-1) + \cdots + \hat{\boldsymbol{A}}_i(m)\hat{\boldsymbol{\theta}}(N-n) \qquad i = 1, 2, \cdots, m$$

置

$$\theta^*(N+1)=(\hat{\theta}_1(N+1),\hat{\theta}_2(N+2),\cdots,\hat{\theta}_m(N+1))^\tau$$

则同样可得出向前两步的预报公式

$$\hat{\theta}_i^*(N+2)=\mathbf{A}_i(0)\hat{\boldsymbol{\theta}}^*(N+1)+\mathbf{A}_i(1)\hat{\boldsymbol{\theta}}(N)+\cdots+\boldsymbol{A}_i(n)\hat{\boldsymbol{\theta}}(N-n+1)\quad i=1,\ 2,\ \cdots,\ m$$

一般地，我们用

$$\hat{\boldsymbol{\theta}}^*(N+j)=(\hat{\theta}_1^*(N+j),\hat{\theta}_2^*(N+j),\cdots,\hat{\theta}_m^*(N+j))^\tau$$

表示时变参量 $\hat{\boldsymbol{\theta}}(k)$ 的向前 j 步预报值，则我们可以依次地得到向前 h 步的预报算法

$$\begin{aligned}\hat{\theta}_i^*(N+h)=&\hat{\boldsymbol{A}}_i(0)\hat{\boldsymbol{\theta}}^*(N+h-1)\\&+\hat{\boldsymbol{A}}_i(1)\hat{\boldsymbol{\theta}}^*(N+h-2)+\cdots\\&i=1,\ 2,\cdots,\ m\end{aligned}\tag{5.10}$$

其中 $\hat{\boldsymbol{A}}_i(0),\hat{\boldsymbol{A}}_i(1),\cdots;\hat{\boldsymbol{A}}_i(n)$ 表示 $\mathbf{A}_i(0),\mathbf{A}_i(1),\cdots,\mathbf{A}_i(n)$ 的估值.

如果 $\mathbf{A}_i(0),\mathbf{A}_i(1),\cdots,\ \mathbf{A}_i(n)$ 仍为时变的，则我们可以引出次一层时间序列，继续进行分析，如此等等.

2. 按分量处理的途径

$\hat{\boldsymbol{\theta}}(k)$ 的各分量之间是相互关联的. 但当它们自身独特的变化规律比较明显时，我们可以考虑把这些分量分开处理，即按分量处理. 一般地说，这将比统一处理的途径简单些. 按分量处理的方法很多，其中有的虽然纯属经验性的，但在实际应用中确是简单而有效的. 这里，仅介绍几种常用的方法.

(1) AR 模型法　这是按分量处理的一般性方法. 我们对 $\hat{\boldsymbol{\theta}}(k)$ 的各分量 $\hat{\theta}_i(k)(i=1,2,\cdots,m)$ 分别建立它们的 AR 模型

$$\begin{aligned}\hat{\theta}_i(k+1)=&\alpha_0\hat{\theta}_i(k)+\alpha_1\hat{\theta}_i(k-1)+\cdots\\&+\alpha_{n_i}\hat{\theta}_i(k-n_i)+e_i(k+1)\\&i=1,\ 2,\cdots,\ m\end{aligned}\tag{5.11}$$

其中 $\{e_i(k)\}$ 是不相关的随机变量序列（通常希望它是零均值的），n_i 是第 i 个模型的阶数. 我们称模型 (5.10) 为第二层 AR

模型. 未知参数 $\alpha_1, \alpha_2, \cdots, \alpha_{n_i}$ 可按前章所介绍的方法来确定. 如果经检验,它们已是非时变的了,则所求的预报公式是

$$\hat{\theta}_i^*(N+1) = \hat{\alpha}_0\hat{\theta}_i(N) + \hat{\alpha}_1\hat{\theta}_i(N-1) + \cdots + \hat{\alpha}_{n_i}\hat{\theta}_i(N-n_i)$$
$$\hat{\theta}_i^*(N+2) = \hat{\alpha}_0\hat{\theta}_i^*(N+1) + \hat{\alpha}_1\hat{\theta}_i(N) + \cdots + \hat{\alpha}_{n_i}\hat{\theta}_i(N-n_i+1)$$
$$\vdots$$

其中 $\hat{\alpha}_0, \hat{\alpha}_1, \cdots, \hat{\alpha}_{n_i}$ 表示 $\alpha_0, \alpha_1, \cdots, \alpha_{n_i}$ 的估值. $\hat{\theta}_i^*(N+h)$, $h \geqslant 1$ 表示 $\hat{\theta}_i(k)$ 的向前 h 步的预报值.

如果 $\alpha_0, \alpha_1, \cdots, \alpha_{n_i}$ 仍为时变的,则我们可以引出第三层时间序列,如此等等.

(2) **定常增量法** 以第 i 个分量 $\hat{\theta}_i(k)$ 为例. 如果对于一切 k 皆有

$$\hat{\theta}_i(k+1) - \hat{\theta}_i(k) = \Delta\theta_i + e_i(k) \qquad i = 1, 2, \cdots, m$$

其中 $\Delta\theta_i$ 是与 k 无关的常数, $\{e_i(k)\}$ 是零均值的白噪声,并具有足够小的方差. 则我们有如下的预报公式

$$\hat{\theta}_i^*(N+h) = \hat{\theta}_i(N) + h\Delta\theta_i \qquad i = 1, 2, \cdots, m \tag{5.12}$$

$\hat{\theta}_i^*(N+h)$, $h \geqslant 1$ 表示 $\hat{\theta}_i(k)$ 的向前 h 步的预报值.

(3) **周期增量法** 我们仍以第 i 个分量 $\hat{\theta}_i(k)$ 为例. 如果它存在着某种周期性的变化,例如它满足

$$\hat{\theta}_i(k+j) - \hat{\theta}_i(k+j-1) = \Delta a_i(j) + e_i(k+z)$$
$$j = 1, 2, \cdots, T$$

而

$$\hat{\theta}_i(k+T+1) - \hat{\theta}_i(k+T) = \Delta a_i(1) + e_i(k+T+1)$$

其中 T 是正整数, $\{e_i(k)\}$ 是零均值的白噪声. 则我们说序列 $\{\hat{\theta}_i(k)\}$ 具有周期增量的变化方式. 此时 $\hat{\theta}(k)$ 的预报公式可以表示成

$$\hat{\theta}_i^*(N+h) = \hat{\theta}_i(N) + p(\Delta a_i(1) + \Delta a_i(2) + \cdots + \Delta a_i(p)) + \Delta a_i(1) + \cdots + \Delta a_i(q) \tag{5.13}$$

其中 p 满足

$$h = pT + q \qquad q < T$$

(4) **常数因子法** 仍以 $\hat{\theta}(k)$ 的第 i 个分量 $\hat{\theta}_i(k)$ 为例. 如

果有常数 q_i，使得

$$\hat{\theta}_i(N+h)=\hat{\theta}_i(N)q_i^h+e_i(N+h)$$

其中 $\{e_i(N+h)\}$ 是零均值方差足够小的随机变量序列. 则 $\hat{\theta}_i(k)$ 的向前 h 步的预报公式是

$$\hat{\theta}_i^*(N+h)=q_i^h\hat{\theta}_i(N) \tag{5.14}$$

这种方法，称为常数因子法. 实质上它相当于 AR 模型法阶数是 1 的情形.

(5) *曲线增量法* 这是定常增量法的推广. 其向前 h 步的预报公式是

$$\hat{\theta}_i^*(N+h)=\hat{\theta}_i(N)+h^{p_i}\Delta\theta_i \qquad i=1,2,\cdots,m$$

$\Delta\theta_i$ 是不依赖于时间的常数. p_i 取某个适当的实数为值. 当 $p_i=1$ 时，就是定常增量法的情形.

6. *其他方法* 关于高层时间序列的预报，还可能有许多不同的方法. 例如，多项式拟合法、样条函数法、直接按其变化规律而得出的直接描述方法(如周期变化法)等等. 在这里就不一一介绍了.

§5.4 预报结果的误差分析

1. 线性情形

考虑系统

$$y_k=\boldsymbol{\phi}(k)^\tau\boldsymbol{\theta}(k)+e_k$$

其中 $\theta(k)$ 是 m 维时变参数，$\boldsymbol{\phi}(k)$ 是 m 维向量值函数，它依赖于 k 时刻及其以前的输入和 $k-1$ 时刻及其以前的输出. 不失一般性，假定这个系统是单输出的. 所要估计的预报误差是

$$\varepsilon(N+h,\hat{\boldsymbol{\theta}}^*(N+h))=y_{N+h}-\hat{y}(N+h)$$

其中 h 是预报的步长. 以 AR 模型为例

$$y_k+a_1(k)y_{k-1}+\cdots+a_m(k)y_{k-m}=e_k$$

此时有

$$\boldsymbol{\phi}(k)=[-y_{k-1},\cdots,-y_{k-m}]^\tau$$

$$\hat{\theta}(k) = [a_1(k), \cdots, a_m(k)]^\tau$$

向前一步的预报公式是

$$y(N+1) = \phi(N+1)^\tau \theta^*(N+1)$$

$$\phi(N+1) = [-y_N, -y_{N-1}, \cdots, -y_{N+1-m}]^\tau$$

向前两步的预报公式是

$$\hat{y}(N+2) = \hat{\phi}(N+2)^\tau \hat{\theta}^*(N+2)$$

$$\hat{\phi}(N+2) = [-\hat{y}(N+1), -y_N, \cdots, -y_{N+2-m}]^\tau$$

向前 h 步的预报公式是

$$\hat{y}(N+h) = \hat{\phi}(N+h)^\tau \hat{\theta}^*(N+h)$$

$$\hat{\phi}(N+h) = [-\hat{y}(N+h-1), -\hat{y}(N+h-2), \cdots]^\tau$$

不难得出

$$\|\phi(N+h) - \hat{\phi}(N+h)\|$$

$$\leqslant \sqrt{m \wedge (h-1)} \max_{1 \vee (h-m) \leqslant i \leqslant h-1} \{|Y_{N+i} - \hat{y}(N+i)|\} \quad \text{a.s}$$

由于预报误差是

$$\varepsilon(N+h, \hat{\theta}^*(N+h)) = \phi(N+h)^\tau \theta(N+h) - \hat{\phi}(N+h)^\tau \hat{\theta}^*(N+h) + e_{N+h}$$

噪声 $e(N+h)$ 是不可避免的，所以误差的可变部分是

$$\phi(N+h)^\tau \theta(N+h) - \hat{\phi}(N+h) \hat{\theta}^*(N+h).$$

这时有

$$\begin{aligned}
&|\phi(N+h)^\tau \theta(N+h) - \hat{\phi}(N+h)^\tau \hat{\theta}^*(N+h)| \\
&\leqslant |\hat{\phi}(N+h)^\tau \tilde{\theta}(N+h)| + |(\phi(N+h) \\
&\quad - \hat{\phi}(N+h))^\tau \hat{\theta}^*(N+h)| \\
&\leqslant \|\phi(N+h)\| \|\tilde{\theta}(N+h)\| \\
&\quad + \|\hat{\theta}^*(N+h)\| \sqrt{m \wedge (h-1)} \\
&\quad \times \max_{1 \vee (h-m) \leqslant i \leqslant h-1} \{|y_{N+i} - \hat{y}(N+i)|\} \quad \text{a.s}
\end{aligned}$$

其中 $\tilde{\theta}(N+h) = \theta(N+h) - \hat{\theta}^*(N+h)$.

2. 非线性情形

不失一般性，仍考虑单输出系统

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + e_k$$

其中

$$\mathbf{Y}_{k-1} = \{y_0, y_1, \cdots, y_{k-1}\}$$

$$\mathbf{U}_k = \{u(0), u(1), \cdots, u(k)\}$$

如果包含 y_k 的预报估计和 $u(k)$ 的预计值时，我们就把相应的 $\mathbf{Y}_{k-1}$ 改写成 $\mathbf{Y}^*_{k-1}$,$\mathbf{U}_k$ 改写成 $\mathbf{U}^*_k$. 在以下的讨论中,记号 $\mathbf{Y}_{k-1}$, $\mathbf{Y}^*_{k-1}$, $\mathbf{U}_k$, $\mathbf{U}^*_k$ 既表示相应的数据集合，又表示向量. 所以记号 $\nabla_{\mathbf{y}_{k-1}} f[\mathbf{Y}_{k-1}, \cdots]$有意义. 在预计的输入已知的情况下,我们有

$$\hat{y}_{N+h} = f[\mathbf{Y}^*_{N+h-1}, \mathbf{U}^*_{N+h}, \hat{\boldsymbol{\theta}}^*(N+h), N+h]$$

$$y_{N+h} = f[\mathbf{Y}_{N+h-1}, \mathbf{U}^*_{N+h}, \boldsymbol{\theta}(N+h), N+h] + e_{N+h}$$

如果相应的梯度都存在,则预报误差的可变部分满足

$$\begin{aligned}
& f[\mathbf{Y}_{N+h-1}, \mathbf{U}^*_{N+h}, \boldsymbol{\theta}(N+h), N+h] \\
& \quad - f[\mathbf{Y}^*_{N+h-1}, \mathbf{U}^*_{N+h}, \hat{\boldsymbol{\theta}}^*(N+h), N+h] \\
& = f[\mathbf{Y}_{N+h-1}, U^*_{N+h}, \boldsymbol{\theta}(N+h), N+h] \\
& \quad - f[\mathbf{Y}^*_{N+h-1}, \mathbf{U}^*_{N+h}, \hat{\boldsymbol{\theta}}^*(N+h), N+h] \\
& \quad + f[\mathbf{Y}^*_{N+h-1}, \mathbf{U}^*_{N+h}, \boldsymbol{\theta}(N+h), N+h] \\
& \quad - f[\mathbf{Y}_{N+h-1}, \mathbf{U}^*_{N+h}, \hat{\boldsymbol{\theta}}^*(N+h), N+h] \\
& = \nabla^*_{\mathbf{y}_{N+h-1}} f[\overline{\mathbf{Y}^*_{N+h-1}}, \mathbf{U}^*_{N+h}, \boldsymbol{\theta}(N+h), \\
& \quad N+h]^\tau(\mathbf{Y}_{N+h-1} - \mathbf{Y}^*_{N+h-1}) \\
& \quad + \nabla_{\theta^*(N+h)} f[\mathbf{Y}^*_{N+h-1}, \mathbf{U}^*_{N+h}, \\
& \quad \overline{\boldsymbol{\theta}^*(N+h)}, N+h]^\tau \tilde{\boldsymbol{\theta}}(N+h)
\end{aligned}$$

其中 $\overline{\mathbf{Y}^*_{N+h-1}}$ 表示 $\mathbf{Y}^*_{N+h-1}$ 与 $\mathbf{Y}_{N+h-1}$ 的连线上的一点; $\overline{\boldsymbol{\theta}^*(N+h)}$ 表示 $\hat{\boldsymbol{\theta}}^*(N+h)$ 与 $\boldsymbol{\theta}(N+h)$ 的连线上的一点;而 $\tilde{\boldsymbol{\theta}}(N+h) = \boldsymbol{\theta}(N+h) - \hat{\boldsymbol{\theta}}^*(N+h)$. 如果有常数 C,使得

$$\|\nabla_{\mathbf{y}^*} f[\mathbf{Y}^*, \mathbf{U}^*_{N+h}, \boldsymbol{\theta}(N+h), N+h]\| < C \quad \text{a.s}$$

$$\|\nabla_{\theta} f[\mathbf{Y}_{N+h-1}, \mathbf{U}^*_{N+h}, \boldsymbol{\theta}, N+h]\| < C \quad \text{a.s}$$

则有

$$\begin{aligned}
& |f[\mathbf{Y}_{N+h-1}, \mathbf{U}^*_{N+h}, \boldsymbol{\theta}(N+h), N+h] \\
& \quad - f[\mathbf{Y}^*_{N+h-1}, \mathbf{U}^*_{N+h}, \hat{\boldsymbol{\theta}}^*(N+h), N+h]|
\end{aligned}$$

$$\leqslant hC \max_{1\leqslant i\leqslant h-1} |Y_{N+i} - \hat{Y}_{N+i}| + C\|\tilde{\theta}(N+h)\| \quad \text{a.s}$$

无论是线性情形还是非线性情形，经上述分析皆表明，向前 h 步的预报误差，依赖于时变参数的预报误差 $\tilde{\theta}(N+h)$ 和以前各步的预报误差 $Y_{N+i} - Y^*_{N+i}$, $i=1,2,\cdots,h-1$. 由于对参量 $\theta(k)$ 进行了预报，所以预报精度能有较大的提高。

这里，我们所进行的只是对预报误差的粗估计。当然，还需要精估计，但在这里就不进行了。

如果我们的预报是依据具有随机时变参数而无附加噪声的模型进行的，只需对上述分析稍加改动就行了。

§5.5 相关噪声干扰下的多层递阶预报方法

我们考虑具有 MA 噪声的系统

$$y_k + a_1 y_{k-1} + \cdots + a_n y_{k-n} = b_0 u(k) + b_1 u(k-1) + \cdots + b_m u(k-m) + e_k + c_1 e_{k-1}, + \cdots + c_p e_{k-p} \tag{5.15}$$

其中 y_k 是一维输出，$u(k)$ 是一维输入，e_k 是零均值的白噪声。

置

$$\phi(k)^\tau = [-y_{k-1},\cdots,-y_{k-n}, u(k), u(k-1),\cdots, u(k-m), e_{k-1},\cdots e_{k-p}\}$$

$$\theta^\tau = [a_1,\cdots,a_n, b_0,\cdots,b_n, c_1,\cdots,c_p]$$

则模型(5.14)可以写成

$$y_k = \phi(k)^\tau\theta + e_k \tag{5.16}$$

但白噪声 $e_{k-1}, e_{k-2}, \cdots, e_{k-p}$ 是不能直接观测的，所以考虑用

$$\hat{\varepsilon}(k) = y_k - \hat{\phi}(k)^\tau\hat{\theta}(k-1) \tag{5.17}$$

来代替它们，其中

$$\hat{\phi}(k)^\tau = [-y_{k-1},\cdots,-y_{k-n}, u(k), u(k-1),\cdots u(k-m), \hat{\varepsilon}(k-1),\cdots,\hat{\varepsilon}(k-p)] \tag{5.18}$$

可见 $\hat{\varepsilon}(k)$ 和 $\hat{\phi}(k)$ 可以交互地被确定。自然在这里我们假定了 θ 是时变的。其跟踪估值可由下式确定

$$\hat{\boldsymbol{\theta}}(k)=\hat{\boldsymbol{\theta}}_1(k)+\frac{1}{\|\hat{\boldsymbol{\phi}}(k)\|^2}\hat{\boldsymbol{\phi}}(k)\{y_k-\hat{\boldsymbol{\phi}}(k)\hat{\boldsymbol{\theta}}_1(k)\}$$

$$\hat{\boldsymbol{\theta}}_1(k)=\hat{\boldsymbol{\theta}}_1(k-1)+M(k)\hat{\varepsilon}(k)$$

$$M(k)=\frac{p(k-1)\hat{\boldsymbol{\phi}}(k)}{\lambda+\hat{\boldsymbol{\phi}}(k)^\tau p(k-1)\hat{\boldsymbol{\phi}}(k)}$$

$$P(k)=\frac{1}{\lambda}[I-M(k)\hat{\boldsymbol{\phi}}(k)^\tau]P(k-1)$$

其中 λ 是遗忘因子.

设应用上述的算法所得到的关于 $\boldsymbol{\theta}(k)$ 的跟踪估值是

$$\hat{\boldsymbol{\theta}}(1),\hat{\boldsymbol{\theta}}(2),\cdots,\hat{\boldsymbol{\theta}}(N)$$

N 是观测数据的组数. 则我们可以把这些估值当作 $\{\boldsymbol{\theta}(k)\}$ 的观测数据而对它进行统计分析. 应用时间序列的多层分析方法，最终我们得出关于 $\boldsymbol{\theta}(k)$ 的向前 h 步的预报估值

$$\hat{\boldsymbol{\theta}}^*(N+1),\hat{\boldsymbol{\theta}}^*(N+2),\cdots,\hat{\boldsymbol{\theta}}^*(N+h)$$

依据这些估值,可得出关于 y_k 的向前 h 步的预报算法.

1. 直接方法

向前一步的预报算法是

$$\hat{y}_{N+1}=\hat{\boldsymbol{\phi}}(N+1)^\tau\hat{\boldsymbol{\theta}}^*(N+1)$$

其中

$$\hat{\boldsymbol{\phi}}(N+1)^\tau=[-y_N,\cdots,-y(N-n+1),\hat{u}(N+1),\\ u(N),\cdots,u(N-m+1),\hat{\varepsilon}(N),\cdots,\hat{\varepsilon}(N-P+1)]$$

$\hat{u}(N+1)$ 表示预计的输入值.

向前两步的预报算法是

$$\hat{y}_{N+2}=\hat{\boldsymbol{\phi}}^*(N+2)^\tau\hat{\boldsymbol{\theta}}^*(N+2)$$

其中

$$\boldsymbol{\phi}^*(N+2)^\tau=[-\hat{y}_{N+1},-y_N,\cdots,-y_{N-n+2},\hat{u}(N+2),\hat{u}(N+1),\\ u(N),\cdots,u(N-m+2),\hat{\varepsilon}^*(N+1),\\ \hat{\varepsilon}(N),\cdots,\hat{\varepsilon}(N-p+2)]$$

$\hat{u}(N+1)$, $\hat{u}(N+2)$ 都是预计的输入值,而

$$\hat{\varepsilon}^*(N+1) = \hat{y}_{N+1} - \boldsymbol{\phi}(N+1)^\tau \hat{\boldsymbol{\theta}}(N)$$

如此继续下去，逐步的可得到向前 h 步的预报算法。

2. 奥斯特隆姆方法

这里，我们只需把在第三章中所介绍的奥斯特隆姆方法稍加改进就行了。为此引入记号

$$\hat{\boldsymbol{\theta}}^*(N+k)^\tau = [\hat{a}_1(N+k), \cdots, \hat{a}_n(N+k), \hat{b}_0(N+k), \hat{b}_1(N+k), \cdots, \hat{b}_m(N+k), \hat{c}_1(N+k), \cdots, \hat{c}_p(N+k)]$$

$$k = 1, 2, \cdots, h$$

$$\hat{A}_{N+k}(q^{-1}) = 1 + \hat{a}_1(N+k)q^{-1} + \cdots + \hat{a}_n(N+k)q^{-n}$$

$$\hat{B}_{N+k}(q^{-1}) = \hat{b}_0(N+k) + \hat{b}_1(N+k)q^{-1} + \cdots + \hat{b}_m(N+k)q^{-m}$$

$$\hat{C}_{N+k}(q^{-1}) = 1 + \hat{c}_1(N+k)q^{-1} + \cdots + \hat{c}_p(N+k)q^{-p}$$

于是可以把非自适应情形下的奥斯特隆姆方法推广如下：

向前 k 步的自适应预报算法是

$$\hat{y}_{N+k} = \hat{G}_{N+k}(q^{-1})\hat{C}_{N+k}^{-1}(q^{-1})y_N + \hat{F}_{N+k}(q^{-1})\hat{B}_{N+k}(q^{-1})\hat{C}_{N+k}^{-1}(q^{-1})\hat{u}(N+k)$$

$k = 1, 2, \cdots, h$. 其中 $\hat{u}(N+k)$ 是预计的输入值。而 $\hat{G}_{N+k}(q^{-1})$ 和 $\hat{F}_{N+k}(q^{-1})$ 满足

$$\hat{A}_{N+k}^{-1}(q^{-1})\hat{C}_{N+k}(q^{-1}) = \hat{F}_{N+k}(q^{-1}) + q^{-k}\hat{G}_{N+k}(q^{-1})\hat{A}_{N+k}^{-1}(q^{-1})$$

$\hat{F}_{N+k}(q^{-1})$ 是 q^{-1} 的 $k-1$ 次多项式

$$\hat{F}_{N+k}(q^{-1}) = 1 + f_1(N+k)q^{-1} + \cdots + f_{k-1}(N+k)q^{-(k-1)}$$

$\hat{G}_{N+k}(q^{-1})$ 是 q^{-1} 的 $n-1$ 次多项式

$$\hat{G}_{N+k}(q^{-1}) = g_0(N+k) + g_1(N+k)q^{-1} + \cdots + g_{n-1}(N+k)q^{-(n-1)}$$

3. 计算实例

这里，我们给出一个应用奥斯特隆姆方法进行预报的实际例子。考虑如下的系统：

$$(1 + a_k q^{-1})y_k = b_k u(k-1) + (1 + c_k q^{-1})e_k$$

对于这个系统，我们有

$$A_k(q^{-1}) = 1 + a_k q^{-1} \quad B_k(q^{-1}) = b_k q^{-1}$$

$$C_k(q^{-1}) = 1 + c_k q^{-1}$$

设应用参数跟踪算法，依据 y_k 和 $u(k)$ 的观测数据，得出了关于时变参数 a_k，b_k 和 c_k 的估值序列如表 5.1.

表 5.1

k	1	2	3	4	5	6	7	8	9	10
$\hat{a}_k$	0.856	0.812	0.770	0.73	0.696	0.667	0.63	0.591	0.581	0.593
$\hat{b}_k$	0.761	0.724	0.684	0.654	0.618	0.584	0.555	0.534	0.506	0.479
$\hat{c}_k$	0.665	0.658	0.652	0.645	0.639	0.633	0.626	0.620	0.614	0.607

由上述的估值序列，不难得出关于各时变参数的预报估值公式

$$\hat{a}_{10+h} = (0.95)^h \hat{a}_{10} \qquad \hat{b}_{10+h} = (0.95)^h \hat{b}_{10}$$

$$\hat{c}_{10+h} = (0.99)^h \hat{c}_{10}$$

应用这些公式得出时变参数的预报估值如表 5.2.

表 5.2

h	1	2	3	4	5
$\hat{a}_{10+h}$	0.512	0.486	0.462	0.439	0.417
$\hat{b}_{10+h}$	0.455	0.432	0.411	0.390	0.371
$\hat{c}_{10+h}$	0.601	0.595	0.589	0.583	0.571

以下我们用"常规的自适应"奥斯特隆姆预报方法和多层递阶奥斯特隆姆预报方法，分别计算 y_k 的预报估值，并对它们进行比较。为此用公式

$$\hat{A}_k(q^{-1})\hat{C}_k(q^{-1}) = \hat{F}_k(q^{-1}) + q^{-h}\hat{G}_k(q^{-1})\hat{A}_k^{-1}(q^{-1})$$

确定多项式 $\hat{G}_k(q^{-1})$ 和 $F_k(q^{-1})$，我们有

$$\hat{G}_k(q^{-1}) = (\hat{c}_k - \hat{a}_k)\hat{a}_k^{h-1}$$

$$\hat{F}_k(q^{-1}) = 1 + (\hat{c}_k - \hat{a}_k)q^{-1} + (\hat{c}_k - \hat{a}_k)\hat{a}_k q^{-2} + \cdots$$

$$+ (\hat{c}_k - \hat{a}_k)\hat{a}_k^{h-2} q^{-h+1}$$

于是得出向前 h 步的"常规自适应"奥斯特隆姆预报算法

$$\hat{y}_{10+h} = \frac{(\hat{c}_{10} - \hat{a}_{10})\hat{a}_{10}^{h-1}}{1 + \hat{c}_{10}q^{-1}} y_{10} + \frac{\hat{F}_{10}(q^{-1})\hat{b}_0}{1 + \hat{c}_{10}q^{-1}} \hat{u}_{10+h-1} \quad (5.19)$$

其中

$$\hat{a}_{10} = 0.593 \quad \hat{b}_0 = 0.429 \quad \hat{c}_{10} = 0.607$$

$\hat{u}(10 + h - 1)$ 表示 $10 + h - 1$ 时刻的预计的输入值.

同样可以得出向前 h 步的多层递阶-奥斯特隆姆预报算法是

$$\hat{y}^*_{10+h} = \frac{(\hat{c}_{10+h} - \hat{a}_{10+h})\hat{a}_{10+h}^{h-1}}{1 + c_{10+h}q^{-1}} y_{10}$$

$$+ \frac{\hat{F}_{10+h}(q^{-1})\hat{b}_{10+h}}{1 + \hat{c}_{10+h}q^{-1}} \hat{u}_{10+h-1} \quad (5.20)$$

其中

$$\hat{F}_{10+h}(q^{-1}) = 1 + (\hat{c}_{10+h} + \hat{a}_{10+h})q^{-1} + (\hat{c}_{10} + \hat{a}_{10+h})\hat{a}_{10+h}q^{-2}$$
$$+ \cdots + (\hat{c}_{10+h} - \hat{a}_{10+h})\hat{a}_{10+h}^{h-2}q^{-h+1}$$

为了对算法的优劣进行比较,我们保留了部分观测数据,见表5.3.

表 5.3

k	10	11	12	13	14	15
$u(k)$	4.6	4.8	4.2	5.1	5.5	
y_k	3.5	0.327	2.1	0.86	1.81	1.4

应用算法(5.19)和(5.20)分别的求出预报估值 $\hat{y}_{10+h}$ 和 $\hat{y}^*_{10+h}$ $(h = 1, 2, 3, 4, 5)$ 见表 5.4.

其中 $y = y_{10+h}$ 是 y_k 的实测值. $\hat{y}^* = \hat{y}^*_{10+h}$ 是用多层递阶奥斯特隆姆方法所得出的预报值. $\hat{y} = \hat{y}_{10+h}$ 是用常规自适应奥斯特隆姆方法所得出的预报值.

比较上述结果可以看出,应用多层递阶奥斯特隆姆方法所得出的预报结果,其误差明显减小。

表 5.4

h	1	2	3	4	5
y	0.327	2.1	0.86	1.81	1.4
$\hat{y}^*$	0.301	2.29	0.835	1.948	1.41
$\hat{y}$	0.162	2.26	0.704	2.071	1.45

§5.6 预报问题的特殊性

前面曾指出以控制系统设计为目的所建立的数学模型，即所谓控制模型，经过适当改变后可以作为以预报为目的的数学模型，即预报模型. 但是预报问题和控制问题确有一定的差别，预报问题有着某种特殊性. 如果忽略了这种特殊性，用处理控制问题的某些观点和方法来处理预报问题，必然会使预报效果受到影响.

控制问题，特别是具有反馈环节的控制系统，要求其数学模型对可能出现的系统状态有较广泛的适应性，即要求模型具有"相空间"的某种大范围的特征. 然而预报问题的解决，基本上是依据系统状态及与之有关的某些量的过去和现在的一组观测值来推断它的未来. 此时，我们所涉及的仅仅是系统状态的一条轨线(实现). 这就是说，预报模型不需要有"相空间"的大范围特征. 但从另一方面看，预报问题总是要涉及一定的"时间尺度"，特别是"长期"预报，其时间尺度就更大了. 所以，可以说，"相空间"中的局部性和"时间尺度"的大范围性，是预报模型所具有的特殊性，从而对于预报模型的辨识和进行预报，也应该有相应的特殊方法.

选择恰当的辨识准则和预报准则，是解决辨识与预报问题的重要步骤. 为了不使预报效果受到不应有的影响，预报模型的辨识准则和进行预报时所用的准则，应该具有某种一致性即统一性.

预报及预报模型辨识所用的准则，应与对预报的精度的要求相匹配. 下面的例子说明，最小二乘辨识和最小方差预报不能实现预报误差任意小.

考虑向前一步的线性预报模型

$$y_k + a_1(k)y_{k-1} + \cdots + a_n(k)y_{k-n} = b_1(k)u(k-1) + \cdots + b_m(k)u(k-m) \qquad (5.21)$$

其中 y_k 是被预报量，$u(k)$ 是预报因子。不妨设它们的维数都是1. $a_1(k),\cdots,a_n(k),b_1(k),\cdots,b_m(k)$ 是未知的时变参数.如果置

$$\boldsymbol{\varphi}(k) = [-y_{k-1},\cdots,-y_{k-n},u(k-1),\cdots,u(k-m)]^\tau$$

$$\boldsymbol{\theta}(k) = [a_1(k),\cdots,a_n(k),b_1(k),\cdots,b_m(k)]^\tau$$

则模型(5.21)可以写成

$$y_k = \boldsymbol{\varphi}(k)^\tau\boldsymbol{\theta}(k)$$

应用递推最小二乘法确定参数 $\boldsymbol{\theta}(k)$ 的估值 $\hat{\boldsymbol{\theta}}(k)$，具体地有

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{P(k-1)}{\lambda + \boldsymbol{\varphi}(k)^\tau P(k-1)\boldsymbol{\varphi}(k)} \times \boldsymbol{\varphi}(k)\{y_k - \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1)\}$$

$$P(k) = \frac{1}{\lambda}\left[I - \frac{P(k-1)\boldsymbol{\varphi}(k)}{\lambda + \boldsymbol{\varphi}(k)^\tau P(k-1)\boldsymbol{\varphi}(k)}\boldsymbol{\varphi}(k)^\tau\right]P(k-1)$$

λ 是遗忘因子。在此基础上，可以得出向前一步的预报算法如下

$$\hat{y}(k|k-1) = \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1) \qquad (5.22)$$

由于原系统(5.21)是无噪声的，所以上述的预报算法可以广义地理解为，是在最小方差意义下所得出的。

应用这种算法所得出的预报误差是

$$y_k - \hat{y}(k|k-1) = \boldsymbol{\varphi}(k)^\tau\boldsymbol{\theta}(k) - \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1) = \boldsymbol{\varphi}(k)^\tau(\boldsymbol{\theta}(k) - \hat{\boldsymbol{\theta}}(k-1))$$

由于 $\boldsymbol{\theta}(k)$ 是时变的，所以不能使得对一切 k 皆有 $|\boldsymbol{\varphi}(k)^\tau(\boldsymbol{\theta}(k) - \hat{\boldsymbol{\theta}}(k-1))|$ 任意小。

上述例子说明，用处理控制问题的某些方法直接来处理预报问题，有时不能得到所希望的结果。所以，应该寻求处理预报问题的特殊方法。

§ 5.7 一致小误差预报

首先引进"具有一致小误差的预报算法"的概念。

设 P 是对于动态系统 S 的一种预报算法，y_k 是被预报量，如果对于任何 $\varepsilon > 0$ 皆可以用算法 P 求得一个预报估值 $\hat{y}_\varepsilon(k)$，使得

$$|y_k - \hat{y}_\varepsilon(k)| < \varepsilon \qquad \text{a.s } \forall k$$

就说 P 是具有一致小误差的预报算法。

可以说明，多层递阶预报算法，在一定条件下是一种具有一致小误差的预报算法。对此我们分两种情形来讨论。

1. 线性情形

考虑向前一步的预报模型

$$y_k = \boldsymbol{\varphi}(k)^\tau \boldsymbol{\theta}(k) \tag{5.23}$$

其中

$$\boldsymbol{\varphi}(k) = [-y_{k-1}, \cdots, -y_{k-n}, u(k-1), \cdots, u(k-n)]^\tau$$

$$\boldsymbol{\theta}(k) = [a_1(k), \cdots, a_n(k), b_1(k), \cdots, b_m(k)]^\tau$$

特别要指出的是 $\boldsymbol{\theta}(k)$ 可以是随机向量，所以模型 (5.23) 具有一定的普遍性。例如，只要置

$$\boldsymbol{\theta}(k) = \boldsymbol{\eta}(k) + \frac{1}{\|\boldsymbol{\varphi}(k)\|^2} \boldsymbol{\varphi}(k) e_k$$

就可以把模型

$$y_k = \boldsymbol{\varphi}(k)^\tau \boldsymbol{\eta}(k) + e_k$$

化为(5.23)式的形式。

我们可以证明，依据模型(5.23)所得出的多层递阶预报算法，在一定的条件下是一种具有一致小误差的预报算法。

事实上，不妨设所用的多层递阶预报算法的层数是 2，其第一层预报模型是

$$\hat{y}(k|k-1)=\boldsymbol{\varphi}(k)^{\tau}\boldsymbol{\theta}(k|k-1) \tag{5.24}$$

进一步假定第二层预报模型所得的预报估值 $\hat{\theta}(k|k-1)$ 满足

$$\|\hat{\boldsymbol{\theta}}(k)-\hat{\theta}(k|k-1)\|<\varepsilon \qquad \text{a.s } \forall k \tag{5.25}$$

其中 $\hat{\boldsymbol{\theta}}(k)$ 是由满足后验残差一致小准则的算法

$$\hat{\theta}(k)=\hat{\boldsymbol{\theta}}(k-1)+\frac{1}{\|\boldsymbol{\varphi}(k)\|^2}\boldsymbol{\varphi}(k)\{y_k-\boldsymbol{\varphi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k-1)\} \tag{5.26}$$

所确定的关于 $\boldsymbol{\theta}(k)$ 的估值．并且有 $c>0$ 使

$$\|\boldsymbol{\varphi}(k)\|<c \qquad \text{a.s } \forall k$$

则我们可以得出

$$|y_k-\hat{y}(k|k-1)|<c\varepsilon \qquad \text{a.s } \forall k$$

这是由于

$$y_k-\hat{y}(k|k-1)=\boldsymbol{\varphi}(k)^{\tau}\boldsymbol{\theta}(k)-\boldsymbol{\varphi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k|k-1)$$

但由算法(5.26)有

$$\varepsilon(k,\hat{\boldsymbol{\theta}}(k))=y_k-\boldsymbol{\varphi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k)=0$$

即

$$\boldsymbol{\varphi}(k)^{\tau}\boldsymbol{\theta}(k)=y_k=\boldsymbol{\varphi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k)$$

所以

$$\begin{aligned}y_k-\hat{y}(k|k-1)&=\boldsymbol{\varphi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k)-\boldsymbol{\varphi}(k)^{\tau}\hat{\boldsymbol{\theta}}(k|k-1)\\&=\boldsymbol{\varphi}(k)^{\tau}(\hat{\boldsymbol{\theta}}(k)-\hat{\boldsymbol{\theta}}(k|k-1)\end{aligned}$$

故有

$$|y_k-\hat{y}(k|k-1)|\leqslant\|\boldsymbol{\varphi}(k)\|\|\hat{\boldsymbol{\theta}}(k)-\hat{\boldsymbol{\theta}}(k|k-1)\|<c\varepsilon \quad \text{a.s } \forall k$$

2. 非线性情形

仍考虑向前一步的多层递阶预报问题．设非线性向前一步的预报模型为

$$y_k=f[\mathbf{Y}_{k-1},\mathbf{U}_k,\boldsymbol{\theta}(k),k] \tag{5.27}$$

其中 $\boldsymbol{\theta}(k)$ 是随机时变参数．它的估值 $\hat{\boldsymbol{\theta}}(k)$ 由下式递推地确定

$$\hat{\boldsymbol{\theta}}(k)=\hat{\boldsymbol{\theta}}(k-1)+\frac{\delta}{\|\nabla_{\boldsymbol{\theta}(k-1)}f[k,\hat{\boldsymbol{\theta}}(k-1)]\|^2}$$

$$\times \nabla_{\theta(k-1)} f[k, \boldsymbol{\theta}(k-1)]$$

$$\times \{y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k-1), k]\} \quad (5.28)$$

应用定理 4.1，不难说明在适当条件下，对于任何的 $\varepsilon > 0$，皆有 δ 存在，使得对于这样的 δ，由算法(5.28)所确定的估值 $\hat{\boldsymbol{\theta}}(k)$ 满足

$$|y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k]| < \varepsilon \qquad \text{a.s } \forall k$$

与线性的情形一样，可以证明：

依据模型 (5.27) 所得出的多层递阶预报算法，在适当的条件下，是一种具有一致小误差的预报算法。

为说明上述结论，不妨假定多层递阶预报算法的层数为 2. 第一层的预报公式是

$$\hat{y}(k|k-1) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k|k-1), k]$$

进一步假定第二层预报模型所得的预报估值 $\hat{\boldsymbol{\theta}}(k|k-1)$ 满足

$$\|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k|k-1)\| < \varepsilon \qquad \text{a.s } \forall k$$

则我们有

$$\begin{aligned} y_k - \hat{y}(k|k-1) &= y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k|k-1), k] \\ &= y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k] \\ &\quad + f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k] - f[\mathbf{Y}_{k-1}, \\ &\quad \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k|k-1), k] \end{aligned}$$

然而

$$|y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k]| < \varepsilon \qquad \text{a.s } \forall k$$

$$f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k] - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k|k-1), k]$$

$$= \nabla_{\theta} f[k, \overline{\boldsymbol{\theta}(k)}]^{\tau}[\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k|k-1)]$$

其中

$$\nabla_{\theta} f[k, \overline{\boldsymbol{\theta}(k)}] = \frac{\partial}{\partial \theta} f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta, k]|_{\theta=\overline{\theta(k)}}$$

$\overline{\boldsymbol{\theta}(k)}$ 是由 $\hat{\boldsymbol{\theta}}(k)$ 到 $\hat{\boldsymbol{\theta}}(k|k-1)$ 的“连线”上的一个适当的点。所以我们有

$$\begin{aligned} |y_k - \hat{y}(k|k-1)| &\leqslant |y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k]| \\ &\quad + \|\nabla_{\theta} f[k, \overline{\boldsymbol{\theta}(k)}]\| \|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k|k-1)\| < \varepsilon \\ &\quad + \|\nabla_{\theta} f[k, \overline{\boldsymbol{\theta}(k)}]\| \varepsilon \end{aligned}$$

由此可见,只要在 θ 的适当的区域中,一致地有

$$\|\nabla_{\theta} f[k,\theta]\| < c \qquad \text{a.s } \forall k$$

c 为大于零的常数。于是有

$$|y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k|k-1), k]| < (1+c)\varepsilon$$

这就是我们需要说明的。

上述讨论说明,只要有时变参数的预报估值算法,使得对于任意的 $\varepsilon > 0$, 皆可找到 $\hat{\theta}(k|k-1)$, 满足

$$\|\hat{\theta}(k) - \hat{\theta}(k|k-1)\| < \varepsilon \qquad \text{a.s } \forall k$$

那么在适当条件下,由多层递阶预报算法所得出的预报估值,就可能具有充分一致小的预报误差。

§ 5.8 预报模型参数估值的初值选择

在多层递阶预报方法的应用中，我们经常考虑的是如下的向前一步的预报模型

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] + V(k) \tag{5.29}$$

其中 y_k 是一维输出，$u(k)$ 是输入，$\theta(k)$ 是 m 维的（时变）参数，f 是关于 $\theta(k)$ 可微的函数，$V(k)$ 是随机噪声，它与 $f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k]$ 相互独立，而

$$\mathbf{Y}_{k-1} = \{y_0, y_1, \cdots, y_{k-1}\}$$

$$\mathbf{U}_k = \{u(0), u(1), \cdots, u(k)\}$$

在预报算法中,所采用的参数估值公式是

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\delta}{\|\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\|^2} \times \nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)]\{y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k-1), k]\} \tag{5.30}$$

其中

$$\nabla_{\theta(k-1)} f[k, \hat{\theta}(k-1)] = \frac{\partial}{\partial \theta} f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta, k]\Big|_{\theta=\theta(k-1)}$$

δ 是适当的常数。

特别重要的情形是预报模型具有下述的形式

$$y_k = \varphi(k)^\tau \theta(k) + v_k \tag{5.31}$$

其中 $\varphi(k)$ 是由 $\mathbf{Y}_{k-1}$, $\mathbf{U}_k$ 的某些元素组成的向量，此时算法(5.30)变为

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\delta}{\|\varphi(k)\|^2}\varphi(k)\{y_k - \varphi(k)^\tau \hat{\theta}(k-1)\} \tag{5.32}$$

并恒取 $\delta = 1$.

递推算法(5.30)和(5.32)所得出的估值,一般会受到初值选择的影响. 但怎样选取这种初值为好、初值的选择对预报精度有什么影响等,这些显然是需要考虑的问题.

如果用 $\hat{\theta}(k|k-1)$ 表示 $\theta(k)$ 的预报估值,则多层递阶方法的基本预报公式是

$$\hat{y}(k|k-1) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k|k-1), k] \tag{5.33}$$

或

$$\hat{y}(k|k-1) = \varphi(k)^\tau \hat{\theta}(k|k-1) \tag{5.34}$$

由于参数估值是用公式(5.30)或(5.32)来确定的，而 $\hat{\theta}(k|k-1)$ 是依赖于这组估值序列求得的，所以它依赖于估值的初值 $\hat{\theta}(0)$. 即

$$\hat{\theta}(k|k-1) = \hat{\theta}(k|k-1, \hat{\theta}(0))$$

从而预报算法(5.33)和(5.34)应该写成

$$\hat{y}(k|k-1) = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}[k|k-1, \hat{\theta}(0)], k]$$

和

$$\hat{y}(k|k-1) = \varphi(k)^\tau \hat{\theta}[k|k-1, \hat{\theta}(0)]$$

自然地,我们称

$$\varepsilon(k, \hat{\theta}(0)) = y_k - \hat{y}(k|k-1)$$

为预报误差.

可以说明,多层递阶预报算法,在一定的意义下，其预报精度(误差)不受 $\hat{\theta}(0)$ 的选取的影响. 对此,我们分两种情形来讨论:

1. 非线性情形

定理 5.1 对于系统(5.29),无论其噪声 $v(k)$ 的统计特性如何,如果有常数 c 使得对一切 k 皆有

$$|y_k| < c \quad \text{a.s}$$

对一切 k, y_k 和 $u(k)$, 一切 $\boldsymbol{\beta} \in R^m$ 皆有

$$|f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\beta}, k]| < c \quad \text{a.s}$$

$$\left|\frac{\partial}{\partial \beta} f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\beta}, k]\right| < c \quad \text{a.s}$$

设 $\{\hat{\boldsymbol{\theta}}(k)\}$ 由估值公式(5.30)所确定,并对 $\delta > 0$ 一致地有

$$\lim_{k\to\infty} \frac{\nabla_{\theta(k-1)} f[k, \boldsymbol{\theta}(\overline{k})]^\tau \nabla_{\theta(k-1)} f[k, \boldsymbol{\theta}(k-1)]}{\|\nabla_{\theta(k-1)} f[k, \hat{\boldsymbol{\theta}}(k-1)]\|^2} = \mu > 0 \quad \text{a.s}$$

其中 $\overline{\theta(k)}$ 由下式决定

$$f[k, \hat{\boldsymbol{\theta}}(k)] = f[k, \hat{\boldsymbol{\theta}}(k-1)] + \nabla_{\theta(k-1)} f[k, \overline{\theta(k)}]^\tau \boldsymbol{\theta}(k)$$

$$\boldsymbol{\theta}(k) = \hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k-1) \quad f[k, \boldsymbol{\theta}] = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \boldsymbol{\theta}, k]$$

则对于任何的 $\varepsilon > 0$, 只要参数 $\boldsymbol{\theta}(k)$ 的预报值 $\hat{\boldsymbol{\theta}}(k|k-1)$ 满足

$$\|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k|k-1)\| < \varepsilon \quad \text{a.s} \ \forall k$$

无论算法的初值 $\hat{\boldsymbol{\theta}}(0)$ 是怎样选取的,对于适当大的 k 都有

$$|\varepsilon(k, \hat{\boldsymbol{\theta}}(0))| = |y_k - \hat{y}(k|k-1)| < M \cdot \varepsilon \quad \text{a.s}$$

此处M为常数。

证明 由于定理 4.1的条件完全得到满足,所以在算法 (5.30)中,必有 $\delta > 0$ 和 $N > 0$, 使得由它所得出的 $\hat{\boldsymbol{\theta}}(k)$ 满足

$$|y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k]| < \varepsilon \quad \text{a.s} \quad k \geqslant N$$

并且与初值 $\hat{\theta}(0)$ 的选择无关,而

$$\begin{aligned}
\boldsymbol{\varepsilon}(k, \hat{\theta}(0)) &= y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k|k-1), k] \\
&= y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k] \\
&\quad + f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k] - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k|k-1), k] \\
&= y(k) - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\boldsymbol{\theta}}(k), k]
\end{aligned}$$

$$+\nabla_{\beta} f[k, \beta]^{\tau}(\theta(k) - \theta(k|k-1))$$

其中

$$\nabla_{\beta} f[k, \beta] = \frac{\partial}{\partial \beta} f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \beta, k]|_{\beta=\bar{\beta}}$$

而 是由微分中值定理所确定的。

所以无论如何选择 $\hat{\theta}(0)$，我们都有

$$|\varepsilon(k, \hat{\theta}(0))| \leqslant |y_k - f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \hat{\theta}(k), k]|$$
$$+ \|\nabla_{\beta} f[k, \bar{\beta}]\| \cdot \|\hat{\theta}(k) - \hat{\theta}(k|k-1)\|$$
$$< \varepsilon + \varepsilon \cdot c = (1+c)\varepsilon \qquad \text{a.s} \quad k \geqslant N$$

于是可见只要取 $M = 1 + c$ 即可。

2. 线性情形

由于对形如

$$y_k = \phi(k)^{\tau}\beta(k) + V_k$$

的模型，只要置

$$\theta(k) = \beta(k) + \frac{1}{\|\varphi(k)\|^2}\varphi(k)V_k$$

就可以化成如下的形式

$$y_k = \varphi(k)^{\tau}\theta(k) \tag{5.35}$$

其中 $\theta(k)$ 是随机时变参数。所以只要考虑形如(5.35)式的系统就够了。

定理 5.2 对于系统(5.35)，如果对一切 k 皆有

$$\|\varphi(k)\| < c \qquad \text{a.s}$$

这里 c 是常数，而且 $\theta(k)$ 的估值 $\hat{\theta}(k)$ 是由下述算法确定的：

$$\hat{\theta}(k) = \theta(k-1) + \frac{1}{\|\varphi(k)\|^2}\varphi(k)\{y_k - \varphi(k)^{\tau}\hat{\theta}(k-1)\} \tag{5.36}$$

则不管算法的初值 $\hat{\theta}(0)$ 怎样选取，对于任何的 $\varepsilon > 0$，只要参数预报估值 $\hat{\theta}(k|k-1)$ 满足

$$\|\hat{\theta}(k) - \hat{\theta}(k|k-1)\| < \varepsilon/c \qquad \text{a.s} \quad \forall k$$

就有

$$|y_k - \hat{y}(k|k-1)| < \varepsilon \qquad \text{a.s} \quad \forall k$$

证明 对于算法(5.36)，显然无论怎样选择初值 $\hat{\theta}(0)$，都有

$$\begin{aligned}\boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k) &= \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1) + \frac{\boldsymbol{\varphi}(k)^\tau}{\|\boldsymbol{\varphi}(k)\|^2} \\ &\quad \times \boldsymbol{\varphi}(k)\{y_k - \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1)\} \\ &= \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1) + y_k - \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1) = y_k\end{aligned}$$

所以

$$\begin{aligned}y_k - \hat{y}(k|k-1) &= \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k) - \boldsymbol{\varphi}(k)^\tau\hat{\boldsymbol{\theta}}(k|k-1) \\ &= \boldsymbol{\varphi}(k)^\tau(\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k|k-1))\end{aligned}$$

故对一切 k 皆有

$$\begin{aligned}|y_k - \hat{y}(k|k-1)| &\leqslant \|\boldsymbol{\varphi}(k)\| \cdot \|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k|k-1)\| \\ &< c \cdot \frac{\varepsilon}{c} = \varepsilon \qquad \text{a.s}\end{aligned}$$

■

从上述两条定理及其证明过程可以看出，多层递阶预报算法的“精度”，依赖于时变参数 $\theta(k)$ 的估值的预报算法的精度，而与 $\theta(k)$ 的估值算法的初值的取法无关.

§5.9 非线性模型线性化的一种途径

众所周知，古典的线性化方法，往往要借助于泰勒(Taylor)展开，所以它难免有局部性的特征. 这种特征，对于预报问题的解决很不利. 因为预报问题希望就“时间尺度”而言，具有一定大范围特性. 所以，自然希望能够找到一种从时间尺度而言，具有大范围特征的线性化方法.

考虑到求解预报问题的基本依据是，系统状态及与之有关的某些量的过去和现在的一组观测值，即我们所涉及的仅仅是系统状态的一条轨线(实现). 所以，我们所寻求的线性化方法，允许从“相空间”(所有的“实现”形成的空间)的观点来看，具有某种局部性.

下面所介绍的，就是具有上述特征的一种线性化途径。

不失一般性，我们来考虑如下的非线性系统

$$y_k = f[\mathbf{Y}_{k-1}, \mathbf{U}_k, \theta(k), k] \tag{5.37}$$

其中 y_k 是一维输出，$u(k)$ 是一维输入，$\theta(k)$ 是 m 维随机时变参数。而

$$\mathbf{Y}_{k-1} = \{y_0, y_1, \cdots, y_{k-1}\}$$
$$\mathbf{U}_k = \{u(0), u(1), \cdots, u(k)\}$$

我们来说明，在一定的意义下，模型(5.37)可被下述形式的一族线性模型所代替

$$y_k = \alpha_1(k)y_{k-1} + \alpha_2(k)y_{k-2} + \cdots + \alpha_p(k)y_{k-p} + \beta_0(k)u(k) + \beta_1(k)u(k-1) + \cdots + \beta_q(k)u(k-q) \tag{5.38}$$

其中 $\alpha_1(k), \alpha_2(k), \cdots, \alpha_p(k), \beta_0(k), \beta_1(k), \cdots, \beta_q(k)$ 是一组随机时变参数。

为说明上述结论，先引入下述的概念。

设 S 是一个动态系统，y_k 是它的输出，$u(k)$ 是它的输入，则称每一组观测数据

$$\{(y_k, u(k))\} \qquad k = 0, 1, 2, \cdots$$

为 S 的一条"轨线"。

定义 设 S_1 是形如(5.37)式的一个非线性模型，$\{S_2\}$ 是形如(5.38)式的一族线性模型，如果对于 S_1 的任何一条轨线 $[Y_1]$，皆有 $\{S_2\}$ 中的一个模型，它的一条轨线 $[Y_2]$ 与 $[Y_1]$ 完全相同，就说 S_1 可嵌入到 $[S_2]$ 之中。

下述定理是关于线性化的主要结论。

定理 5.3 设动态系统满足非线性模型(5.37)，则必存在一族具有随机时变参数的线性形式的模型 $\{m_a \ a \in A\}$:

$$y_k = \alpha_1(k)y_{k-1} + \alpha_2(k)y_{k-2} + \cdots + \alpha_p(k)y_{k-p} + \beta_0(k)u(k) + \beta_1(k)u(k-1) + \cdots + \beta_q(k)u(k-q) \tag{5.39}$$

A 是某个适当的指标集合，使得模型 (5.37) 可嵌入模型族 $\{m_a, a \in A\}$ 之中，这里 p 和 q 的确定可具有一定的任意性。

证明 设 $\{y_k^*, u^*(k)\}$ $k=0, 1, 2, \cdots$是模型(5.37)的任一条轨线. 所需证明的是在模型族 $\{m_a, a\in A\}$ 中必有一个模型 m_a^*,它满足

$$\begin{aligned} y_k^* = \alpha_1^*(k)y_{k-1}^* + \cdots + \alpha_p^*(k)y_{k-p}^* \\ + \beta_0^*(k)u^*(k) + \beta_1^*(k)u^*(k-1) + \cdots \\ + \beta_q^*(k)u^*(k-q) \end{aligned} \tag{5.40}$$

如果置

$$\boldsymbol{\varphi}(k)^\tau = [y_{k-1}, y_{k-2}, \cdots, y_{k-p}, u(k), u(k-1), \cdots, u(k-q)]$$

$$\boldsymbol{\varphi}^*(k)^\tau = [y_{k-1}^*, y_{k-2}^*, \cdots, y_{k-p}^*, u^*(k), u^*(k-1), \cdots, u^*(k-q)]$$

$$\boldsymbol{\eta}(k)^\tau = [\alpha_1(k), \alpha_2(k), \cdots, \alpha_p(k), \beta_0(k), \beta_1(k), \cdots, \beta_q(k)]$$

$$\boldsymbol{\eta}^*(k)^\tau = [\alpha_1^*(k), \alpha_2^*(k), \cdots, \alpha_p^*(k), \beta_0^*(k), \beta_1^*(k), \cdots, \beta_q^*(k)]$$

则(5.39)式和(5.40)式可分别写成

$$y_k = \boldsymbol{\varphi}(k)^\tau \boldsymbol{\eta}(k)$$

和

$$y_k^* = \boldsymbol{\varphi}^*(k)^\tau \boldsymbol{\eta}^*(k)$$

由于 $\{\mathbf{y}_k^*, \mathbf{u}^*(k)\}$ $k=0, 1, 2, \cdots$, 是已知的,当然它满足

$$y_k^* = f[\mathbf{Y}_{k-1}^*, \mathbf{U}_k^*, \boldsymbol{\theta}(k), k]$$

其中

$$\mathbf{Y}_{k-1}^* = \{y_0^*, y_1^*, \cdots, y_{k-1}^*\}$$

$$\mathbf{U}_k^* = \{u^*(0), u^*(1), \cdots, u^*(k)\}$$

为证明我们的定理,只需说明,必有 $\eta^*(k)$ 满足

$$\boldsymbol{\varphi}^*(k)^\tau \boldsymbol{\eta}^*(k) = y_k^*$$

就行了. 这几乎是显然的,只要取

$$\boldsymbol{\eta}^*(k) = \boldsymbol{\eta}_0 + \frac{1}{\|\boldsymbol{\varphi}^*(k)\|^2}\boldsymbol{\varphi}^*(k)\{y_k^* - \boldsymbol{\varphi}^*(k)^\tau \boldsymbol{\eta}_0\}$$

就够了. 其中 $\boldsymbol{\eta}_0$ 是任意取定的向量. 在实际应用中, $\eta^*(k)$ 往

往要递推地确定，故一般的取 $\eta_0 = \eta^*(k-1)$.

由于对如上选取的 $\eta^*(k)$ 有

$$
\begin{aligned}
\varphi^*(k)^\tau \eta^*(k) &= \varphi^*(k)^\tau \eta_0 + \varphi^*(k)^\tau \frac{1}{\|\varphi^*(k)\|^2} \\
&\quad \times \varphi^*(k)\{y_k^* - \varphi^*(k)^\tau \eta_0\} \\
&= \varphi^*(k)^\tau \eta_0 + y_k^* - \varphi^*(k)^\tau \eta_0 = y_k^*
\end{aligned}
$$

所以模型(5.37)可嵌入模型族 $\{m_a, a \in A\}$ 之中. ■

上述定理既说明了非线性模型在一定意义下是可以进行大范围线性化的(就时间域而言)，也给出了构造这种线性化模型的方法.

根据上述定理，在处理一般的动态系统的预报问题时，我们仅考虑形如(5.38)的预报模型就够了. 但由于这种模型具有随机时变参数，所以所采用的预报方法，应该是多层递阶方法.

第六章　天气与农业系统的预报

这里，我们简单地介绍多层递阶方法在长期天气预报和农业系统预报中的应用情况．可以看出，多层递阶方法在解决长期天气预报及农业预报这类困难问题中，其效果是比较理想的．

§6.1　天气系统及其预报问题

大气系统是一个非常复杂的大系统，它由许多互相联系、互相制约的小天气系统（子系统）组成．能量的变化是大气系统变化的根本原因．大气系统的大范围变化是许许多多子系统变化的综合结果．

从整个地球来看，由外部输入的能量基本上来源于太阳．而大气的能量获得主要是来自其下垫面，如海洋、陆地等等．下垫面能量的变化与天气系统的许多因素有关，但其中云量变化是主要因素之一．

天气形势是以各种气象要素来表征的．从系统论的观点来看，各种气象要素可以认为是相应系统的输出．而来自大气系统外部的能量，可以看成是系统的输入．无论是"大的"天气系统还是"小的"天气系统，都是多输入-多输出的复杂系统．另一方面，从天气系统本身的结构来看，又有着一种随时间变化的特征，这种变化，有的缓慢、有的迅速．所以天气系统是一个时变的复杂动态系统．为进行天气过程的预报，就必须清楚地揭示天气系统的动态变化规律．因为预报问题的基本点，就是要依据事物自身的发展变化规律来揭示、推断它的未来．为揭示这一规律，常用而有效的方法是正确地建立描述系统的动态数学模型．从系统理论的观点来看，这就是系统建模或辨识问题．

如果已清楚地了解了系统变化规律的机理，我们可以根据机理来建立系统的模型．这样的模型称之为机理模型，它反映了系统内部因素之间的联系．因而有着明显的物理意义．例如，为大家所熟知的大气热力-流体动力学方程组，其中包括牛顿第二定律方程、热力学第一定律方程、连续方程、状态方程、水份守恒方程等等．这些方程总起来，可以大体上从大气系统的内部结构来描述大气的运动变化状态． 假如这组方程表示了所有因素之间的关系，而且问题的初始和边界条件为已知，则从理论上讲，它能够将预报任何未来时刻的大气状况．不幸的是，实际上能够应用的方程组，虽然描述了多少接近于真实的一个大气模型，但是它们总是不完全的．况且，大气的初始条件也很难准确地确定．虽然可以相信，随着人们对大气过程的知识的增加，会使大气热力-流体动力学方程组逐渐完善．但是，原始数据的不充分将导致预报质量随活动时间周期的增加而迅速下降．另一方面，尽管已经产生了多少接近于真实大气运动的机理模型，然而相应方程组的准确求解几乎是不可能的，因为它具有很复杂的形式．当然，我们可以依赖于数值解，但是前述的大气初始状态的不准确性，仍给我们带来很大的困难，更不幸的是，许多经验证明，从现在熟知的大气热力-流体动力学方程出发，预报大气状态的最大周期仅有一个星期左右．超过这个周期，由解方程组所作的预报与纯随机预报已无太大差别．这无疑给用这种方法进行长期天气预报又带来了新的困难．

上述种种，并不是企图说明，在长期天气预报中，从大气热力-流体动力学方程组出发的途径是行不通的．而仅仅是提出了一系列困难．毫无疑问，克服这些困难的途径是会被人们找到的．事实上，现在有些气象工作者已经做出了一些有益的工作．

建立天气系统模型的另一个途径，就是从系统的外部特征着手，建立它的输入-输出模型，或"纯输出"模型．对于建立这种模型最方便的途径是，把单一的天气要素或某些紧密关联的要素的组合的随时间变化的数据，看成一个一维的或多维的时间序列．

然后用时间序列分析的方法建模，如带有输入（预报因子）的ARMA 模型、AR 模型等等．以这些模型为基础，进行时间序列的多层分析和多层递阶预报．这可能使我们避免某些非线性模型,从而可能避免由于非线性模型所带来的某些复杂现象．

应用时间序列的多层分析方法,将要依据大量的历史资料(数据)．这可以使预报模型的建模过程所依据的信息大大增多,使所得的模型能较好地反映系统的历史演变规律，从而有利于提高模型对于长期预报的适应性．

时间序列模型,因为重点不是依据动态系统的机理而导出的．所以物理意义不明显．因而这种方法与观点，不易为某些气象理论工作者所接受．然而大量的实践资料说明，时间序列的多层分析与多层递阶预报方法,在解决长期天气预报的某些问题时,其效果是良好的．

在这一章中,仅用少数的实例来说明这种情况．

§6.2 关于春季平均气温的长期预报

这里我们所介绍的是，周期增量法在单一气象要素的长期预

表 6.1

年　份	1957	1958	1959	1960	1961	1962	1963	1964	1965
k	1	2	3	4	5	6	7	8	9
y_k(℃)	3.6	2.9	6.5	3.0	4.5	5.0	6.0	5.4	3.7

年　份	1966	1967	1968	1969	1970	1971	1972	1973	1974
k	10	11	12	13	14	15	16	17	18
y_k(℃)	4.3	6.6	7.3	3.6	5.1	4.0	5.9	3.9	5.0

年　份	1975	1976	1977	1978	1979	1980	1981	1982
k	19	20	21	22	23	24	25	26
y_k(℃)	7.6	5.6	5.8	5.7	5.3	4.3	5.2	6.4

报中的应用情况。

我们已得到了某地区历年来春季平均气温的数据如表 6.1。

应用这些春季（3—5 月）平均气温的原始数据，经分析可以确定春季平均气温所满足的动态方程为

$$y_k = \theta_1(k)y_{k-1} + \theta_2(k)y_{k-2} + \theta_3(k)y_{k-3} + e(k) \qquad (6.1)$$

其中 y_k 是时刻 k 的春季平均气温，$\theta_1(k)$，$\theta_2(k)$，$\theta_3(k)$ 皆为时变参数，k 是流动时间，$e(k)$ 是随机噪声。

对于系统(6.1)的时变参数有如下的跟踪公式

$$\begin{pmatrix} \hat{\theta}_1(k) \\ \hat{\theta}_2(k) \\ \hat{\theta}_3(k) \end{pmatrix} = \begin{pmatrix} \hat{\theta}_1(k-1) \\ \hat{\theta}_2(k-1) \\ \hat{\theta}_3(k-1) \end{pmatrix} + \frac{1}{y_{k-1}^2 + y_{k-2}^2 + y_{k-3}^2} \begin{pmatrix} y_{k-1} \\ y_{k-2} \\ y_{k-3} \end{pmatrix}$$

表 6.2

k	$\hat{\theta}_1(k)$	$\hat{\theta}_2(k)$	$\hat{\theta}_3(k)$
4	0.3100	0.1151	0.1809
5	0.4406	0.3981	0.3071
6	0.4209	0.3858	0.2804
7	0.5263	0.4870	0.3437
8	0.3948	0.3774	0.2384
9	0.2817	0.2517	0.1337
10	0.3331	0.3268	0.2171
11	0.5284	0.4948	0.4624
12	0.5261	0.4933	0.4611
13	0.1793	0.1798	0.2568
14	0.2268	0.2760	0.3438
15	0.1903	0.2502	0.2916
16	0.3950	0.5112	0.4759
17	0.1721	0.3601	0.2832
18	0.2354	0.4559	0.3481
19	0.4081	0.5906	0.5519
20	0.2059	0.4575	0.4481
21	0.1533	0.3862	0.4012
22	0.1523	0.3851	0.3512
23	0.1535	0.3861	0.3912

$$\cdot \{y_k - \hat{\theta}_1(k-1)y_{k-1} - \hat{\theta}_2(k-1)y_{k-2} - \hat{\theta}_3(k-1)y_{k-3}\} \tag{6.2}$$

其中 $\hat{\theta}_1(k-1)$，$\hat{\theta}_2(k-1)$，$\hat{\theta}_3(k-1)$ 表示 $\theta_1(k-1)$，$\theta_2(k-1)$，$\theta_3(k-1)$ 的估值.

依据前述数据，并选取时变参数的初值

$\hat{\theta}_1(0) = 0.3100$　　$\hat{\theta}_2(0) = 0.1151$　　$\hat{\theta}_3(0) = 0.1809$

应用时变参数的跟踪公式(6.2)，可得出时变参数的一系列估值如表 6.2.

对上述的关于 $\theta_1(k)$，$\theta_2(k)$ 和 $\theta_3(k)$ 的估值序列进行分析可以发现，它们各自的变化规律，即它们的增量大体上都有以 4 为周期的周期变化特征.

如果从第 24 年开始计算，则这些周期变化的增量值分别表示如下：

关于 $\hat{\theta}_1(k)$ 的周期增量的估值是

$$-0.1121,\ -0.0910,\ +0.0537,\ +0.1560$$

关于 $\hat{\theta}_2(k)$ 的周期增量的估值是

$$-0.1414,\ -0.0757,\ +0.0890,\ +0.1346$$

关于 $\hat{\theta}_3(k)$ 的周期增量的估值是

$$-0.1101,\ -0.0845,\ +0.0784,\ +0.1151$$

应用这些增量可从第 24 年开始，对时变参数 $\theta_1(k)$，$\theta_2(k)$，$\theta_3(k)$ 的值进行预报. 进而又可以对预报量 $y(k)$ 进行预报.

应用多层递阶预报方法所得的时刻 24，25，26（即 80，81，82 年）的时变参数，$y(k)$ 的预报值及其误差等见表 6.3.

表 6.3

k	年份	$\hat{\theta}_1(k)$	$\hat{\theta}_2(k)$	$\hat{\theta}_3(k)$	$\hat{y}_k$(℃)	y_k(℃)	δ	$\left\|\frac{\delta}{y}\right\|$(%)
24	1980	0.0623	0.3105	0.3167	3.9	4.3	−0.4	9.3
25	1981	0.1160	0.3995	0.3951	4.8	5.2	−0.4	7.7
26	1982	0.2720	0.5341	0.5102	6.1	6.4	−0.3	4.7

表中 $\hat{y}_k$ 是预报值，y_k 是实际观测值，且

$$\hat{y}_k = \hat{\theta}_1(k)y_{k-1} + \hat{\theta}_2(k)y_{k-2} + \hat{\theta}_3(k)y_{k-3}$$

$$\delta = \hat{y}_k - y_k$$

为残差，$\left|\frac{\delta}{y}\right|$ 为相对误差，故由以上结果知，预报与实际观测值的平均相对误差为 7%，预报效果较好.

继续应用上述方法，我们预报出今后 18 年某县春季平均气温，见表 6.4.

表 6.4

年　份	k	$\hat{\theta}_1(k)$	$\hat{\theta}_2(k)$	$\hat{\theta}_3(k)$	$\hat{y}_k$
1983	27	0.1599	0.3927	0.4001	4.8
1984	28	0.0689	0.3170	0.3156	4.0
1985	29	0.1226	0.4060	0.3940	5.0
1986	30	0.2786	0.5406	0.5091	6.0
1987	31	0.1665	0.3992	0.3990	4.6
1988	32	0.0755	0.3235	0.3145	3.9
1989	33	0.1292	0.4125	0.3929	4.8
1990	34	0.2852	0.5471	0.5080	5.8
1991	35	0.1731	0.4057	0.3979	4.5
1992	36	0.0821	0.3300	0.3134	3.8
1993	37	0.1358	0.4190	0.3918	4.7
1994	38	0.2918	0.5536	0.5069	5.8
1995	39	0.1797	0.4122	0.3968	4.5
1996	40	0.0887	0.3365	0.3123	3.8
1997	41	0.1424	0.4255	0.3907	4.7
1998	42	0.2984	0.5601	0.5058	5.8
1999	43	0.1863	0.4187	0.3957	4.6
2000	44	0.0953	0.3430	0.3112	3.9

§ 6.3　西太平洋夏季各月份副高面积指数长期预报

在以下两节，我们介绍关于天气形势的长期预报.

西太平洋副热带高压(以下简称副高)在夏季作为影响我国降水和气温分布的一个重要环流系统，国内外气象工作者对它的长

期预报问题进行了大量的研究．这方面的预报方法主要分为两大类：一是从副高本身的变化中寻找规律的预报；另一是用与副高变化密切相关的物理因子，利用各种回归分析方法建立预报方程来做副高的预报．

这里我们应用多层递阶预报方法，进行夏季各月份(6，7，8)副高面积指数的长期预报，取得了良好的效果．

前已指出，对于天气系统这样的时变动态系统，我们可以把关于天气的预报量，称为天气系统的输出变量，而把预报因子称为该系统的输入变量．

表 6.5　西太平洋6月份副高面积指数 $y_1(k)$
(1955～1974 年)有关数据表

k	年份	$y_1(k)$	年份	$u_{11}(k)$	$u_{21}(k)$	$u_{31}(k)$	$u_{41}(k)$	$u_{51}(k)$
1	1955	19	1954	494	0.52	0.74	136	520.6
2	1956	11	1955	502	1.70	0.51	140	507.5
3	1957	17	1956	498	0.80	0.78	141	522.7
4	1958	20	1957	499	1.83	0.54	137	507.7
5	1959	24	1958	504	1.77	0.58	147	511.8
6	1960	24	1959	497	1.23	0.78	142	507.5
7	1961	17	1960	496	0.82	0.76	139	516.2
8	1962	19	1961	499	1.20	0.58	138	520.4
9	1963	21	1962	500	1.75	0.65	139	506.0
10	1964	15	1963	497	1.78	0.60	137	506.6
11	1965	17	1964	501	1.37	0.82	134	508.7
12	1966	26	1965	500	1.38	0.91	135	503.7
13	1967	13	1966	499	0.29	0.81	140	518.5
14	1968	18	1967	507	1.12	0.83	133	511.3
15	1969	28	1968	498	1.52	0.63	137	505.8
16	1970	27	1969	504	0.63	0.90	137	510.1
17	1971	21	1970	501	1.32	0.54	143	510.1
18	1972	17	1971	492	1.18	0.76	136	509.4
19	1973	20	1972	497	1.50	0.66	142	508.6
20	1974	3	1973	497	1.43	0.80	140	508.8

根据长期预报经验，我们发现前期冬季的后 3 个月（12 月，当年 1 月，2 月）的极涡、环流指数、东亚大槽、副高面积指数及鄂霍次克海高压等方面的因子分别与夏季各月份(6，7，8)副高面积指数有较密切的关系。经过对上述各方面的诸因子认真的分析比较和整理剔除，选取前期冬季后三月（12，1，2）的极涡强度（u_{11}，u_{12}，u_{13}），亚洲地区环流经向指数（u_{21}，u_{22}，u_{23}）纬向指数（u_{31}，u_{32}，u_{33}）东亚大槽位置（u_{41}，u_{42}，u_{43}），鄂霍次克海高压（u_{51}，u_{52}，u_{53}），作为预报因子，即为系统的输入变量，而将夏季各月（6,7,8）的副高面积指数作为预报量，即为系统的输出变量（y_1，y_2，y_3）。

表 6.6　西太平洋 7 月份副高面积指数 $y_2(k)$

（1955～1974 年）有关数据表

k	年份	$y_2(k)$	$u_{12}(k)$	$u_{22}(k)$	$u_{32}(k)$	$u_{42}(k)$	$u_{52}(k)$
1	1955	11	502	1.13	0.83	142	510.9
2	1956	7	500	0.93	0.87	138	510.9
3	1957	21	488	1.53	0.55	146	503.6
4	1958	17	503	1.11	0.70	142	509.9
5	1959	23	502	1.20	0.94	138	507.9
6	1960	30	504	0.97	0.56	143	515.9
7	1961	20	500	1.22	0.77	142	510.1
8	1962	22	492	1.35	0.64	142	507.0
9	1963	25	497	0.96	0.67	137	521.2
10	1964	18	489	1.55	0.67	151	507.9
11	1965	14	503	1.28	0.75	140	508.0
12	1966	29	499	1.45	0.74	149	503.4
13	1967	10	491	1.58	0.74	145	498.9
14	1968	18	504	0.61	0.64	143	516.7
15	1969	29	503	1.00	0.50	141	509.9
16	1970	25	503	0.96	0.67	140	520.0
17	1971	17	490	1.53	0.57	140	508.8
18	1972	5	490	1.43	0.62	143	504.6
19	1973	20	495	1.74	0.48	146	501.8
20	1974	10	495	0.80	0.52	144	522.3

1. 预报系统的数学模型

对于输入变量 u_{ij} ($j=1, 2, 3, 4, 5; i=1, 2, 3$). 及输出变量 y_i ($i=1, 2, 3$)，我们已掌握它们的历史数据如表 6.5，6.6，6,7.

由表 6.5—6.7 的数据可确定西太平洋夏季各月份副高面积指数长期预报系统的数学模型为

$$y_i(k)=\varphi_i^{\mathrm{T}}(k)\theta_i(k)+v_i(k) \quad j=1, 2, 3 \tag{6.3}$$

其中 $y_i(k)$ 为一维输出，$v_i(k)$ 为一维白噪声，而

表 6.7 西太平洋8月份副高面积指数 ($y_3(k)$)
(1955～1974 年)有关数据表

k	年份	$y_3(k)$	$u_{13}(k)$	$u_{23}(k)$	$u_{33}(k)$	$u_{43}(k)$	$u_{53}(k)$
1	1955	12	498	0.93	0.63	137	516.8
2	1956	12	496	0.31	0.74	138	520.6
3	1957	14	508	0.60	0.61	138	523.1
4	1958	26	506	1.19	0.78	141	513.2
5	1959	9	488	1.36	0.56	155	509.8
6	1960	24	493	1.85	0.50	143	505.8
7	1961	31	492	1.17	0.57	143	514.1
8	1962	30	496	1.16	0.64	144	509.8
9	1963	28	484	1.45	0.62	143	510.3
10	1964	14	496	0.50	0.72	154	519.8
11	1965	10	494	0.93	0.62	150	516.4
12	1966	26	497	1.89	0.69	141	499.6
13	1967	18	489	1.16	0.62	149	507.7
14	1968	17	502	0.52	0.51	151	525.5
15	1969	23	503	0.85	0.69	141	509.0
16	1970	23	496	0.95	0.58	147	517.5
17	1971	8	500	0.55	0.52	141	525.5
18	1972	18	492	1.22	0.59	136	513.6
19	1973	23	494	0.89	0.71	151	511.4
20	1974	10	506	1.00	0.61	142	513.3

$$\boldsymbol{\varphi}_j^T(k) = (y_j(k-1),\ y_j(k-2),\ u_{1j}(k),\ u_{2j}(k),\ u_{3j}(k),$$
$$u_{4j}(k),\ u_{5j}(k))$$
$$\boldsymbol{\theta}_j^T(k) = (\alpha_{1j}(k),\ \alpha_{2j}(k),\ \beta_{1j}(k),\ \beta_{2j}(k),\ \beta_{3j}(k),\ \beta_{4j}(k),$$
$$\beta_{5j}(k)).$$

系统(6.3)时变参数的跟踪公式如下

$$\hat{\boldsymbol{\theta}}_j(k) = \hat{\boldsymbol{\theta}}_j(k-1) + \frac{\delta}{\|\varphi_j(k)\|^2}\ \boldsymbol{\varphi}_j(k)\{y_j(k) - \boldsymbol{\varphi}_j^T(k)\hat{\boldsymbol{\theta}}_j(k-1)\} \quad j = 1,\ 2,3 \tag{6.4}$$

其中 $\hat{\boldsymbol{\theta}}_j(k-1)$ 表示 $\theta_j(k-1)$ 的估值. δ 可选取为 1.

于是我们首先运用跟踪公式 (6.4) 对系统 (6.3) 的时变参数进行跟踪. 得到时变参数估值序列 $\{\hat{\boldsymbol{\theta}}_j(k)\}$ $(j = 1, 2, 3;\ k = 1, 2, \cdots, N;\ N$ 为观测数据组数). 然后对其进行分析，寻找其规律，建立其所满足的预报模型 (这里可按 $\hat{\boldsymbol{\theta}}_j(k)$ 的各分量分别建立所适合的预报模型). 得到 $\boldsymbol{\theta}_j(k)$ 向前一步的预报估值 $\hat{\boldsymbol{\theta}}_j^*(N+1)$，再可确定 $y_j(k)$ 向前一步的预报公式为

$$\hat{y}_j(N+1|N) = \boldsymbol{\varphi}_j^T(N+1)\hat{\boldsymbol{\theta}}_j^*(N+1) \quad j = 1,\ 2,\ 3 \tag{6.5}$$

其中 $\hat{\boldsymbol{\theta}}_j^*(N+1) = (\hat{\alpha}_{1j}^*(N+1),\ \hat{\alpha}_{2j}^*(N+1),\ \hat{\beta}_{1j}^*(N+1),$
$\hat{\beta}_{2j}^*(N+1),\ \hat{\beta}_{3j}^*(N+1),\ \hat{\beta}_{4j}^*(N+1),$
$\hat{\beta}_{5j}^*(N+1))$

2. 预报结果

我们用多层递阶预报方法，依照掌握的历史数据，逐年对 1975—1982 年进行了 8 年共计 24 个单项的向前一步试验预报. 即逐年补充前期冬季后 3 个月新掌握的数据，运用公式(6.5)向前预报一步.

下面以预报 1975 年 6 月份副高面积指数 (y_1) 为例说明以上预报方法的应用. 设已掌握的新数据如下:

$u_{11}(21) = 494 \qquad u_{21}(21) = 0.74$

$u_{31}(21) = 0.75 \qquad u_{41}(21) = 147$

$u_{51}(21) = 513.7$

表 6.8　时变参数 $\alpha_{11},\alpha_{21},\beta_{11},\beta_{21},\beta_{31},\beta_{41},\beta_{51}$ 的估值表

k	$\hat{\alpha}_{11}(k)$	$\hat{\alpha}_{21}(k)$	$\hat{\beta}_{11}(k)$	$\hat{\beta}_{21}(k)$	$\hat{\beta}_{31}(k)$	$\hat{\beta}_{41}(k)$	$\hat{\beta}_{51}(k)$
3	0.1955	0.1922	−0.0038	0.1997	0.1997	0.1423	−0.0139
4	0.1956	0.1923	−0.0004	0.1997	0.1997	0.1432	−0.0105
5	0.1956	0.1923	0.0004	0.1997	0.1997	0.1435	−0.0069
6	0.1956	0.1923	−0.0002	0.1997	0.1997	0.1433	−0.0102
7	0.1953	0.1920	−0.0069	0.1997	0.1997	0.1414	−0.0172
8	0.1954	0.1921	−0.0036	0.1997	0.1997	0.1423	−0.0137
9	0.1955	0.1922	−0.0012	0.1997	0.1997	0.1430	−0.0113
10	0.1952	0.1920	−0.0074	0.1997	0.1997	0.1413	−0.0176
11	0.1953	0.1921	−0.0042	0.1997	0.1997	0.1421	−0.0144
12	0.1956	0.1924	0.0049	0.1997	0.1997	0.1446	−0.0052
13	0.1949	0.1919	−0.0095	0.1997	0.1997	0.1406	−0.0202
14	0.1951	0.1922	−0.0033	0.1997	0.1997	0.1422	−0.0140
15	0.1954	0.1924	0.0070	0.1998	0.1997	0.1450	−0.0035
16	0.1952	0.1923	0.0034	0.1998	0.1997	0.1440	−0.0072
17	0.1948	0.1919	−0.0048	0.1997	0.1997	0.1417	−0.0155
18	0.1947	0.1918	−0.0064	0.1997	0.1997	0.1413	−0.0171
19	0.1948	0.1919	−0.0026	0.1997	0.1997	0.1424	−0.0132
20	0.1942	0.1914	−0.0182	0.1997	0.1997	0.1380	−0.0292

表 6.9　西太平洋6月份副高面积指数（y_1）(1975～1982 年)有关数据表

k	年份	$y_1(k)$	年份	$u_{11}(k)$	$u_{21}(k)$	$u_{31}(k)$	$u_{41}(k)$	$u_{51}(k)$
21	1975	15	1974	494	0.74	0.75	147	513.7
22	1976	7	1975	491	1.07	0.76	145	517.7
23	1977	21	1976	504	1.13	0.75	136	515.1
24	1978	19	1977	499	1.52	0.64	145	509.8
25	1979	30	1978	497	1.93	0.59	149	502.6
26	1980	33	1979	496	1.59	0.67	139	511.7
27	1981	21	1980	503	1.19	0.67	131	525.0
28	1982	24	1981	497	1.60	0.88	141	501.8

由表 6.5 的历史数据，可取时变参数的初值为：

$$\hat{\alpha}_{11}(0)=0.2 \qquad \hat{\alpha}_{21}(0)=0.19$$
$$\hat{\beta}_{11}(0)=-0.004 \qquad \hat{\beta}_{21}(0)=0.2$$
$$\hat{\beta}_{31}(0)=0.2 \qquad \hat{\beta}_{41}(0)=0.14$$
$$\hat{\beta}_{51}(0)=-0.014$$

然后由跟踪公式 (6.4)，可得到时变参数的一系列估值如表 6.8.

对时变参数估值序列 $\{\hat{\alpha}_{i1}(k)\}\{\hat{\beta}_{l1}(k)\}(i=1, 2;\ l=1, 2, 3, 4, 5)$ 进行分析，可分别根据各自的规律，建立其预报公式。这里，我们对 $\{\hat{\alpha}_{11}(k)\}$，$\{\hat{\alpha}_{21}(k)\}$，$\{\hat{\beta}_{41}(k)\}$ 采用参数的"按分量的AR 模型法"，建立各自的 7 阶 AR 模型；对 $\{\hat{\beta}_{11}(k)\}$，$\{\hat{\beta}_{51}(k)\}$ 采用"分段增量法"；而对简单情形下的 $\{\hat{\beta}_{21}(k)\}$，$\{\hat{\beta}_{31}(k)\}$ 则采用"取均值法"。

于是，我们由时变参数各自的预报公式可得：$\hat{\alpha}_{11}^*(21)=0.1943$，$\hat{\alpha}_{21}^*(21)=0.1921$，$\hat{\beta}_{11}^*(21)=-0.0125$，$\hat{\beta}_{21}^*(21)=0.1997$，$\hat{\beta}_{31}^*(21)=0.1997$，$\hat{\beta}_{41}^*(21)=0.1423$，$\hat{\beta}_{51}^*(21)=-0.0136$。进一步由公式(6.5)，便可得到

$$\begin{aligned}\hat{y}_1(21)=&\ \hat{\alpha}_{11}^*(21)y_1(20)+\hat{\alpha}_{21}^*(21)y_1(19)+\hat{\beta}_{11}^*(21)u_{11}(21)\\&+\hat{\beta}_{21}^*(21)u_{21}(21)+\hat{\beta}_{31}^*(21)u_{31}(21)\\&+\hat{\beta}_{41}^*(21)u_{41}(21)+\hat{\beta}_{51}^*(21)u_{51}(21)\end{aligned}$$

表 6.10 西太平洋7月份副高面积指数 (y_2)

(1975～1982 年)有关数据表

k	年份	$y_2(k)$	$u_{12}(k)$	$u_{22}(k)$	$u_{32}(k)$	$u_{42}(k)$	$u_{52}(k)$
21	1975	14	491	1.17	0.61	147	513.7
22	1976	11	502	0.99	0.78	145	514.2
23	1977	12	508	0.34	0.53	143	521.1
24	1978	6	498	1.33	0.64	140	504.3
25	1979	32	499	1.64	0.70	148	504.4
26	1980	26	505	0.85	0.78	134	517.1
27	1981	25	493	0.64	0.71	153	516.4
28	1982	18	500	1.43	0.73	139	507.1

表 6.11　西太平洋8月副高面积指数（y_3）（1975～1982 年）有关数据表

k	年	$y_3(k)$	$u_{13}(k)$	$u_{23}(k)$	$u_{33}(k)$	$u_{43}(k)$	$u_{53}(k)$
21	1975	13	505	0.99	0.59	145	516.1
22	1976	21	490	1.33	0.66	142	507.5
23	1977	16	500	1.11	0.65	153	513.0
24	1978	20	497	1.24	0.75	152	510.0
25	1979	21	488	1.71	0.65	134	508.7
26	1980	26	505	0.85	0.72	150	512.7
27	1981	18	499	0.78	0.68	142	517.1
28	1982	18	495	1.08	0.58	144	519.5

表 6.12　西太平洋 6 月份副高面积指数（y_1）预报结果表

k	年份	$\hat{\alpha}^*_{11}(k)$	$\hat{\alpha}^*_{21}(k)$	$\hat{\beta}^*_{11}(k)$	$\hat{\beta}^*_{21}(k)$	$\hat{\beta}^*_{31}(k)$	$\hat{\beta}^*_{41}(k)$	$\hat{\beta}^*_{51}(k)$	$\hat{y}_1(k)$	$y_1(k)$	δ_1	$\nabla_1(\%)$	$\hat{y}_1$ 距平值	y_1 距平值	趋势评定
21	1975	0.1943	0.1921	−0.0125	0.1997	0.1997	0.1423	−0.0136	12	15	−3	20	−6	−3	✓
22	1976	0.1942	0.1921	−0.0187	0.1997	0.1997	0.1423	−0.0137	8	7	1	14.3	−10	−11	✓
23	1977	0.1942	0.1921	0.0074	0.1997	0.1997	0.1422	−0.0141	20	21	−1	4.8	2	3	✓
24	1978	0.1940	0.1921	−0.0068	0.1997	0.1997	0.1423	−0.0138	17	19	−2	10.5	−1	1	×
25	1979	0.1943	0.1921	0.0036	0.1997	0.1997	0.1423	−0.0064	28	30	−2	6.7	10	12	✓
26	1980	0.1944	0.1921	0.0076	0.1997	0.1997	0.1424	−0.0025	32	33	−1	3.0	14	15	✓
27	1981	0.1940	0.1921	−0.0025	0.1997	0.1997	0.1424	−0.0129	23	21	2	9.5	5	3	✓
28	1982	0.1941	0.1921	−0.0025	0.1997	0.1997	0.1425	−0.0130	23	24	−1	4.2	5	6	✓

表 6.13　西太平洋 7 月份副高面积指数（y_2）预报结果表

k	年份	$\hat{\alpha}^*_{12}(k)$	$\hat{\alpha}^*_{22}(k)$	$\hat{\beta}^*_{12}(k)$	$\hat{\beta}^*_{22}(k)$	$\hat{\beta}^*_{32}(k)$	$\hat{\beta}^*_{42}(k)$	$\hat{\beta}^*_{52}(k)$	$\hat{y}_2(k)$	$y_2(k)$	δ_2	$\nabla_2(\%)$	$\hat{y}_2$ 矩平值	y_2 矩平值	趋势评定
21	1975	0.1949	0.1951	−0.0107	0.1993	0.1998	0.1368	−0.0182	12	14	−2	14.3	−6	−4	✓
22	1976	0.1949	0.1952	−0.0086	0.1993	0.1998	0.1375	−0.0160	12	11	1	9.1	−6	−7	✓
23	1977	0.1949	0.1952	−0.0099	0.1993	0.1998	0.1371	−0.0174	11	12	−1	8.3	−7	−6	✓
24	1978	0.1949	0.1952	−0.0086	0.1993	0.1998	0.1375	−0.0254	7	6	1	16.7	−11	−12	✓
25	1979	0.1948	0.1951	0.0084	0.1993	0.1998	0.1359	0.0007	27	32	−5	15.6	9	14	✓
26	1980	0.1950	0.1957	0.0023	0.1993	0.1998	0.1432	−0.0067	25	26	−1	3.8	7	8	✓
27	1981	0.1946	0.1956	−0.0046	0.1993	0.1998	0.1413	0.0137	24	25	−1	4	6	7	✓
28	1982	0.1942	0.1951	−0.0048	0.1993	0.1998	0.1391	−0.0115	21	18	3	14.3	4	0	✓

表 6.14　西太平洋 8 月份副高面积指数（y_3）预报结果表

k	年份	$\hat{\alpha}^*_{13}(k)$	$\hat{\alpha}^*_{23}(k)$	$\hat{\beta}^*_{13}(k)$	$\hat{\beta}^*_{23}(k)$	$\hat{\beta}^*_{33}(k)$	$\hat{\beta}^*_{43}(k)$	$\hat{\beta}^*_{53}(k)$	$\hat{y}_3(k)$	$y_3(k)$	δ_3	$\nabla_3(\%)$	$\hat{y}_3$ 矩平值	y_3 矩平值	趋势评定
21	1975	0.1927	0.1945	−0.0053	0.1998	0.1998	0.1433	−0.0217	14	13	1	7.7	−4	−5	✓
22	1976	0.1928	0.1946	−0.0009	0.1998	0.1998	0.1432	−0.0127	18	21	−3	14.3	0	3	✓
23	1977	0.1930	0.1948	−0.0054	0.1998	0.1998	0.1433	−0.0147	19	16	3	18.7	1	−2	×
24	1978	0.1927	0.1946	−0.0056	0.1998	0.1998	0.1432	−0.0126	20	20	0	0	2	2	✓
25	1979	0.1928	0.1948	−0.0055	0.1998	0.1998	0.1432	−0.0106	19	21	−2	9.5	1	3	✓
26	1980	0.1929	0.1949	−0.0020	0.1998	0.1998	0.1438	−0.0093	24	26	−2	7.7	6	8	✓
27	1981	0.1930	0.1949	−0.0051	0.1998	0.1998	0.1433	−0.0123	21	18	3	14.3	3	0	✓
28	1982	0.1926	0.1946	−0.0076	0.1998	0.1998	0.1423	−0.0151	18	18	0	0	0	0	✓

$$
\begin{aligned}
&= 0.1943 \times 3 + 0.1921 \times 20 - 0.0125 \times 494 \\
&\quad + 0.1997 \times 0.74 + 0.1997 \times 0.75 \\
&\quad + 0.1423 \times 147 - 0.0136 \times 513.7 \\
&\approx 12
\end{aligned}
$$

类似地，也可得到 1975～1982 年 y_1, y_2, y_3 的预报结果. 为此，先补充 1975 年以后逐年掌握的有关新数据（见表 6.9～6.11）.

然后，用多层递阶预报方法进行向前一步的预报，其结果及误差分析见表 6.12～6.14.

在表 6.12～6.14 中，$\hat{y}_i(k)$ 为预报值，$y_i(k)$ 为真实值，$\delta_i = \hat{y}_i - y_i$ 为残差. $\nabla_i = \left|\frac{\delta_i}{y_i}\right|$ 为相对误差，$i = 1, 2, 3$. 历年的平均值为：$\bar{y}_1 = 18$, $\bar{y}_2 = 18$, $\bar{y}_3 = 18$，预报的趋势评定中："√"表示正确，"×"表示不正确.

从上述预报结果不难看出，运用多层递阶预报方法来进行夏季各月份副高面积指数的长期预报，由于方法比较适当，再加上选取的输入变量适当，故而预报趋势的准确率可达 91.7%，预报的平均相对误差为 9.6%，这是一般长期预报方法难以达到的.

§6.4 西太平洋夏季各月份副高西伸脊点长期预报

在前节我们曾经运用多层递阶预报方法，建立了夏季西太平洋 6～8 各月份副高面积指数的长期预报模型. 预报的效果比较好.

这里，我们用"多层递阶预报方法"对西太平洋夏季各月份(6～8)副高西伸脊点进行长期预报，建立了副高西伸脊点自适应的长期预报模型. 在预报过程中充分考虑到副高这个动态系统的时变特性. 先对系统的时变参数进行预报，然后再在此基础上对系统的状态（或输出）进行预报. 因而克服了一般迴归预报方法由于固定参数模型而造成的较大预报误差，使预报趋势准确率及

表 6.15　西太平洋副高西伸脊点 1955～1974 历史数据表

k	年份	$y_6(k)$	$y_7(k)$	$y_8(k)$	$y_2(k)$	$y_1(k)$	年份	$y_{12}(k-1)$	$y_{11}(k-1)$
1	1955	120	134	130	155	146	1954	95	150
2	1956	110	115	125	155	155	1955	155	140
3	1957	128	115	134	155	114	1956	150	119
4	1958	125	140	115	124	100	1957	119	129
5	1959	105	110	135	109	115	1958	110	138
6	1960	134	110	133	135	105	1959	105	105
7	1961	113	114	108	155	155	1960	85	90
8	1962	110	110	105	105	155	1961	145	104
9	1963	135	114	105	150	145	1962	105	85
10	1964	110	107	110	135	155	1963	135	105
11	1965	130	113	108	155	105	1964	155	155
12	1966	110	109	103	135	145	1965	109	108
13	1967	130	150	109	155	155	1966	145	120
14	1968	125	144	139	155	155	1967	145	128
15	1969	115	115	125	115	134	1968	155	95
16	1970	120	129	133	135	155	1969	115	140
17	1971	120	119	138	155	155	1970	125	120
18	1972	128	150	128	155	130	1971	139	140
19	1973	114	129	110	99	155	1972	125	105
20	1974	140	135	149	155	155	1973	149	140

预报精度均有明显的提高。

1. 预报系统的数学模型

由于副高西伸脊点受多种因素影响，情况是比较复杂的．我们认为它的自身发展规律就是各种影响因素综合作用的结果．或者说，它本身就是各种影响因素综合作用的产物．所以在这里我们采用了纵横相结合的自适应长期预报模型．即：从纵的角度考虑，由 3 年左右一周期的变化规律，建立预报当月与前 3 年内同月副高西伸脊点之间的适应关系．从横的角度考虑，由前期冬季各月份（11～2 月）西伸脊点与夏季预报各月份（6～8 月）西伸脊点之间比较密切的影响关系．建立起它们之间的相应关系，故令

$y_j(k)$ 表示副高西伸脊点第 k 年 j 月的数值，其中 k 为流动的年份，$j=1, 2, 6, 7, 8, 11, 12$。

现已掌握的1955～1974年副高西伸脊点的历史数据见表6.15。

由以上历史数据，我们可确定夏季各月份(6～8)副高西伸脊点长期预报系统的数学模型为

$$y_j(k)=\boldsymbol{\varphi}_j^{\tau}(k)\boldsymbol{\theta}_j(k)+v_j(k) \quad j=6, 7, 8 \tag{6.6}$$

其中 $y_j(k)$ 为一维输出，$v_j(k)$ 为一维白噪声。而

$$\boldsymbol{\varphi}_j^{\tau}(k)=(y_j(k-1), y_j(k-2), y_j(k-3), y_2(k), y_1(k), y_{12}(k-1), y_{11}(k-1))$$

$$\boldsymbol{\theta}_j^{\tau}(k)=(\alpha_{1j}(k), \alpha_{2j}(k), \alpha_{3j}(k), \beta_{1j}(k), \beta_{2j}(k), \beta_{3j}(k), \beta_{4j}(k))$$

系统(6.6)时变参数的跟踪公式为

$$\hat{\boldsymbol{\theta}}_j(k)=\hat{\boldsymbol{\theta}}_j(k-1)+\frac{\delta}{\|\boldsymbol{\varphi}_j(k)\|^2}\boldsymbol{\varphi}_j(k)\{y_j(k)-\boldsymbol{\varphi}_j^{\tau}(k)\hat{\boldsymbol{\theta}}_j(k-1)\} \quad j=6, 7, 8 \tag{6.7}$$

其中 $\hat{\boldsymbol{\theta}}_j(k-1)$ 表示 $\boldsymbol{\theta}_j(k-1)$ 的估值。我们可取 $\delta=1$。

在实际应用中，往往将(6.7)式写成如下形式

$$\begin{pmatrix}\hat{\alpha}_{1j}(k)\\ \hat{\alpha}_{2j}(k)\\ \hat{\alpha}_{3j}(k)\\ \hat{\beta}_{1j}(k)\\ \hat{\beta}_{2j}(k)\\ \hat{\beta}_{3j}(k)\\ \hat{\beta}_{4j}(k)\end{pmatrix}=\begin{pmatrix}\hat{\alpha}_{1j}(k-1)\\ \hat{\alpha}_{2j}(k-1)\\ \hat{\alpha}_{3j}(k-1)\\ \hat{\beta}_{1j}(k-1)\\ \hat{\beta}_{2j}(k-1)\\ \hat{\beta}_{3j}(k-1)\\ \hat{\beta}_{4j}(k-1)\end{pmatrix}+\delta/[y_j^2(k-1)+y_j^2(k-2)+y_j^2(k-3)+y_2^2(k)+y_1^2(k)+y_{12}^2(k-1)+y_{11}^2(k-1)]$$

$$
\cdot\begin{pmatrix} y_j(k-1) \\ y_j(k-2) \\ y_j(k-3) \\ y_2(k) \\ y_1(k) \\ y_{12}(k-1) \\ y_{11}(k-1) \end{pmatrix}\{y_j(k)-\hat{\alpha}_{1j}(k-1)y_j(k-1)
$$

$$
-\hat{\alpha}_{2j}(k-1)y_j(k-2)-\hat{\alpha}_{3j}(k-1)y_j(k-3)
$$

$$
-\hat{\beta}_{1j}(k-1)y_2(k)-\hat{\beta}_{2j}(k-1)y_1(k)
$$

$$
-\hat{\beta}_{3j}(k-1)y_{12}(k-1)-\hat{\beta}_{4j}(k-1)y_{11}(k-1)\} \quad (6.7)^*
$$

其中 $\hat{\alpha}_{ij}(k-1)$，$\hat{\beta}_{ij}(k-1)$（$i=1,2,3;\ j=1,2,3,4$）分别为 $\alpha_{ij}(k-1)$，$\beta_{ij}(k-1)$ 的估值．显然(6.7)* 式便于我们按 $\hat{\boldsymbol{\theta}}_j(k)$ 的分量分别讨论处理．

由时变参数预报公式可得到向前一步的参数预报估值 $\hat{\boldsymbol{\theta}}_j^*(N+1)$．

进一步可确定 $y_j(k)$ 向前一步的预报公式为

$$
\hat{y}_j(N+1)=\boldsymbol{\varphi}_j^{\tau}(N+1)\hat{\boldsymbol{\theta}}_j^*(N+1) \quad j=6,7,8 \quad (6.8)
$$

其中

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_j^{*\tau}(N+1)=(&\hat{\alpha}_{1j}^*(N+1),\ \hat{\alpha}_{2j}^*(N+1),\ \hat{\alpha}_{3j}^*(N+1),\\
&\hat{\beta}_{1j}^*(N+1),\hat{\beta}_{2j}^*(N+1),\hat{\beta}_{3j}^*(N+1),\\
&\hat{\beta}_{4j}^*(N+1))
\end{aligned}
$$

$\hat{\alpha}_{ij}^*(N+1)$，$\hat{\beta}_{lj}^*(N+1)$ 分别表示 $\alpha_{ij}(N+1)$，$\beta_{lj}(N+1)$ 的估值．$i=1,2,3$；$l=1,2,3,4$．

2. 预报的结果

由"多层递阶预报"方法，依据历史数据，我们逐年对 1975～1982 年进行了 8 年共计 24 个单项的向前一步副高西伸脊点的试验预报．即逐步补充新掌握的数据，运用公式（6.8）向前预报一步．

下面以预报 1975 年 6 月份副高西伸脊点（y_6）为例说明预报

方法的应用（为此先补充已掌握的新数据如下：　$y_6(21)=155$，$y_1(21)=145$、$y_{12}(20)=130$，$y_{11}(20)=130$）.

由表 6.15 的历史数据，可选取时变参数的初值为：$\hat{\alpha}_{16}(0)=0.17$，$\hat{\alpha}_{26}(0)=0.13$，$\hat{\alpha}_{36}(0)=0.08$，$\hat{\beta}_{16}(0)=0.27$，$\hat{\beta}_{26}(0)=0.18$，$\hat{\beta}_{36}(0)=0.13$，$\hat{\beta}_{46}(0)=0.07$. 然后由跟踪公式 (6.7) 或 (6.7)*对时变参数进行跟踪，可得到时变参数的一系列估值，参见表 6.16.

表 6.16　时变参数 α_{16}，α_{26}，α_{36}，β_{16}，β_{26}，β_{36}，β_{46} 的估值表

k	$\hat{\alpha}_{16}(k)$	$\hat{\alpha}_{26}(k)$	$\hat{\alpha}_{36}(k)$	$\hat{\beta}_{16}(k)$	$\hat{\beta}_{26}(k)$	$\hat{\beta}_{35}(k)$	$\hat{\beta}_{45}(k)$
4	0.1711	0.1309	0.0810	0.2710	0.1809	0.1310	0.0711
5	0.1532	0.1126	0.0653	0.2554	0.1666	0.1152	0.0513
6	0.1797	0.1441	0.0975	0.2895	0.1956	0.1417	0.0778
7	0.1474	0.1189	0.0675	0.2522	0.1704	0.1213	0.0561
8	0.1412	0.1115	0.0617	0.2464	0.1618	0.1133	0.0504
9	0.1620	0.1329	0.0871	0.2748	0.1912	0.1332	0.0665
10	0.1304	0.1071	0.0606	0.2432	0.1573	0.1016	0.0420
11	0.1385	0.1171	0.0687	0.2546	0.1687	0.1130	0.0534
12	0.1381	0.1167	0.0683	0.2542	0.1683	0.1126	0.0530
13	0.1432	0.1227	0.0734	0.2614	0.1750	0.1194	0.0586
14	0.1342	0.1151	0.0644	0.2506	0.1643	0.1093	0.0469
15	0.1344	0.1154	0.0646	0.2508	0.1746	0.1096	0.0499
16	0.1407	0.1221	0.0716	0.2582	0.1719	0.1158	0.0575
17	0.1328	0.1146	0.0635	0.2481	0.1618	0.1077	0.0496
18	0.1380	0.1198	0.0684	0.2547	0.1685	0.1137	0.0557
19	0.1465	0.1278	0.0764	0.2613	0.1771	0.1220	0.0627
20	0.1490	0.1305	0.0790	0.2646	0.1804	0.1252	0.0657

对时变参数估值序列 $\{\hat{\alpha}_{i6}(k)\}$ $\{\hat{\beta}_{l6}(k)\}(i=1,2,3;\ l=1,2,3,4)$ 进行分析，并根据其变化规律，可采用“分段增量法”确定其预报公式，即通过分析 $\{\hat{\alpha}_{i6}(k)\}$，$\{\hat{\beta}_{l6}(k)\}$ 不难发现，每个参数估值序列中的增量 $\Delta\hat{\alpha}_{i6}(k)=\hat{\alpha}_{i6}(k)-\hat{\alpha}_{i6}(k-1)$，$\Delta\hat{\beta}_{l6}(k)=\hat{\beta}_{l6}(k)-\hat{\beta}_{l6}(k-1)$ 的变化规律具有一定的周期性，其周期为 7；于是可分段确定 $\Delta\hat{\alpha}_{i6}(k)$ 及 $\Delta\hat{\beta}_{l6}(k)$ 在一个周期变化范围内各段估值，从中排除了个别的特殊情况，适当调整后作为增量预报值 $\Delta\hat{\alpha}_{i6}(k)$ 及 $\Delta\hat{\beta}_{l6}(k)$，这样就可得到时变参数的预报公式为

$\hat{\alpha}^*_{i6}(N+1) = \hat{\alpha}_{i6}(N) + \Delta\hat{\alpha}^*_{i6}(N+1) \quad i = 1, 2, 3$

$\hat{\beta}^*_{l6}(N+1) = \hat{\beta}_{l6}(N) + \Delta\hat{\beta}^*_{l6}(N+1) \quad l = 1, 2, 3, 4$

$\Delta\hat{\alpha}^*_{i6}(k)$，$\Delta\hat{\beta}^*_{l6}(k)$ 的值见表 6.17。

表 6.17 时变参数增量预报值 $\Delta\hat{\alpha}^*_{i6}(k)$，$\Delta\hat{\beta}^*_{l6}(k)$ 表

$(i = 1, 2, 3;\ l = 1, 2, 3, 4)$

k	$\Delta\hat{\alpha}^*_{16}(k)$	$\Delta\hat{\alpha}^*_{26}(k)$	$\Delta\hat{\alpha}^*_{36}(k)$	$\Delta\hat{\beta}^*_{16}(k)$	$\Delta\hat{\beta}^*_{26}(k)$	$\Delta\hat{\beta}^*_{36}(k)$	$\Delta\hat{\beta}^*_{46}(k)$
21,28	−0.0090	−0.0076	−0.0088	−0.0108	−0.0107	−0.0101	−0.0089
22,29	+0.0051	+0.0060	+0.0051	+0.0072	+0.0067	+0.0068	+0.0056
23,30	−0.0179	−0.0183	−0.0157	−0.0156	−0.0143	−0.0158	−0.0198
24,31	+0.0265	+0.0315	+0.0322	+0.0341	+0.0290	+0.0265	+0.0265
25,32	−0.0323	−0.0252	−0.0300	−0.0373	−0.0252	−0.0204	−0.0217
26,33	−0.0062	−0.0074	−0.0057	−0.0058	−0.0086	−0.0080	−0.0057
27,34	+0.0208	+0.0214	+0.0255	+0.0284	+0.0294	+0.0199	+0.0161

于是，由表 6.17 按周期增量法便可得：

$\hat{\alpha}^*_{16}(21) = 0.14$，$\hat{\alpha}^*_{26}(21) = 0.1229$，$\hat{\alpha}^*_{36}(21) = 0.0702$，

$\hat{\beta}^*_{16}(21) = 0.2538$，$\hat{\beta}^*_{26}(21) = 0.1697$，$\hat{\beta}^*_{36}(21) = 0.1151$，

$\hat{\beta}^*_{46}(21) = 0.0568$。再由(6.8)式可得

$$\begin{aligned}\hat{y}_6(21) &= \hat{\alpha}^*_{16}(21)y_6(20) + \hat{\alpha}^*_{26}(21)y_6(19) + \hat{\alpha}^*_{36}(21)y_6(18)\\ &\quad + \hat{\beta}^*_{16}(21)y_2(21) + \hat{\beta}^*_{26}(21)y_1(21)\\ &\quad + \hat{\beta}^*_{36}(21)y_{12}(20) + \hat{\beta}^*_{46}(21)y_{11}(20)\\ &= 0.14 \times 140 + 0.1229 \times 114 + 0.0702 \times 128\\ &\quad + 0.2538 \times 155 + 0.1697 \times 145\\ &\quad + 0.1151 \times 130 + 0.0568 \times 130\\ &\approx 129\end{aligned}$$

类似地，也可得到 1975～1982 年 y_6，y_7，y_8 的预报值。为此，需先补充 1975 年以后逐年新掌握的有关数据，见表 6.18。

经过对相应的参数值进行分析发现，它们也都具有周期增量，所以采用周期增量法对参数进行预报。其周期增量估值见表 6.19。

表 6.18 西太平洋副高西伸脊点 1975～1982 年历史数据表

k	年份	$y_6(k)$	$y_7(k)$	$y_8(k)$	$y_2(k)$	$y_1(k)$	年份	$y_{12}(k-1)$	$y_{11}(k-1)$
21	1975	130	124	135	155	145	1974	130	130
22	1976	140	142	110	155	155	1975	115	155
23	1977	108	120	142	135	145	1976	105	140
24	1978	120	160	110	100	109	1977	115	135
25	1979	104	105	130	100	150	1978	100	135
26	1980	104	115	110	110	130	1979	104	85
27	1981	126	120	115	110	145	1980	145	115
28	1982	110	135	138	115	145	1981	105	105

表 6.19 时变参数增量估值 $\Delta\hat{\alpha}^*_{i7}(k)$, $\Delta\hat{\beta}^*_{l7}(k)$ 周期变化表

($i=1, 2, 3$; $l=1, 2, 3, 4$)

k	$\Delta\hat{\alpha}^*_{17}(k)$	$\Delta\hat{\alpha}^*_{27}(k)$	$\Delta\hat{\alpha}^*_{37}(k)$	$\Delta\hat{\beta}^*_{17}(k)$	$\Delta\hat{\beta}^*_{27}(k)$	$\Delta\hat{\beta}^*_{37}(k)$	$\Delta\hat{\beta}^*_{47}(k)$
21,27	−0.0066	−0.0064	−0.0065	−0.0078	−0.0084	−0.0078	−0.0060
22,28	+0.0177	+0.0168	+0.0178	+0.0212	+0.0165	+0.0171	+0.0170
23,29	−0.0170	−0.0152	−0.0190	−0.0204	−0.0204	−0.0164	−0.0158
24,30	+0.0273	+0.0296	+0.0264	+0.0355	+0.0356	+0.0318	+0.0321
25,31	−0.0241	−0.0251	−0.0183	−0.0192	−0.0260	−0.0260	−0.0159
26,32	+0.0065	+0.0064	+0.0083	+0.0092	+0.0062	+0.0051	+0.0053

利用表 6.20 中的数据对参数进行预报，在此基础上再对预报量 y_6, y_7, y_8 进行预报，其结果见表 6.21～6.23. 于是，逐年运用参数预报公式(6.9),(6.10)(包括表 6.17,6.19,6.20) 及公式(6.8)，便可得到 1975～1982 年预报的结果，见表 6.21. 表中：$\hat{y}_i(k)$ 为预报值，$y_i(k)$ 为真实值，$\delta_i=\hat{y}_i-y_i$ 为残差，$\nabla_i=\left|\frac{\delta_i}{y_i}\right|$ 为相对误差，$i=6, 7, 8$. 历年平均值为 $\bar{y}_6=120$, $\bar{y}_7=122$, $\bar{y}_8=122$，预报趋势评定中的"√"表示正确，"×"表示不正确.

表 6.20　时变参数增量估值 $\Delta\hat{\alpha}^*_{i8}(k)$, $\Delta\hat{\beta}^*_{l8}(k)$ 周期变化表

$(i=1, 2, 3;\ l=1, 2, 3, 4)$

k	$\Delta\hat{\alpha}^*_{18}(k)$	$\Delta\hat{\alpha}^*_{28}(k)$	$\Delta\hat{\alpha}^*_{38}(k)$	$\Delta\hat{\beta}^*_{18}(k)$	$\Delta\hat{\beta}^*_{28}(k)$	$\Delta\hat{\beta}^*_{38}(k)$	$\Delta\hat{\beta}^*_{48}(k)$
21,26	−0.0136	−0.0115	−0.0134	−0.0135	−0.0116	−0.0105	−0.0106
22,27	−0.0342	−0.0348	−0.0296	−0.0319	−0.0270	−0.0219	−0.0231
23,28	+0.0333	+0.0389	+0.0363	+0.0316	+0.0291	+0.0319	+0.0401
24,29	−0.0141	−0.0136	−0.0127	−0.0158	−0.0158	−0.0142	−0.0143
25,30	+0.0254	+0.0240	+0.0251	+0.0361	+0.0361	+0.0338	+0.0298

表 6.21　西太平洋 6 月份副高西伸脊点（y_6）预报结果表

k	年份	$\hat{\alpha}^*_{16}(k)$	$\hat{\alpha}^*_{26}(k)$	$\hat{\alpha}^*_{36}(k)$	$\hat{\beta}^*_{16}(k)$	$\hat{\beta}^*_{26}(k)$	$\hat{\beta}^*_{36}(k)$	$\hat{\beta}^*_{46}(k)$	$\hat{y}_6(k)$	$y_6(k)$	δ_6	∇_6 (%)	$\hat{y}_6$ 矩平值	y_6 矩平值	趋势评定
21	1975	0.1400	0.1229	0.0702	0.2538	0.1697	0.1151	0.0568	129	130	−1	0.8	9	10	✓
22	1976	0.1455	0.1296	0.0762	0.2623	0.1782	0.1240	0.0633	138	140	−2	1.4	18	20	✓
23	1977	0.1300	0.1133	0.0620	0.2485	0.1662	0.1081	0.0468	117	108	9	8.3	−3	−12	✓
24	1978	0.1468	0.1375	0.0823	0.2715	0.1809	0.1297	0.0655	116	120	−4	3.3	−4	0	✓
25	1979	0.1189	0.1197	0.0562	0.2280	0.1570	0.1147	0.0549	92	104	−12	11.5	−28	−16	✓
26	1980	0.1309	0.1256	0.0651	0.2484	0.1626	0.1161	0.0560	101	104	−3	2.9	−19	−16	✓
27	1981	0.1351	0.1512	0.0934	0.2797	0.1971	0.1412	0.0755	132	126	6	4.8	12	6	✓
28	1982	0.1394	0.1363	0.0754	0.2554	0.1766	0.1309	0.0661	115	110	5	4.5	−5	−10	✓

表 6.22　西太平洋 7 月份副高西伸脊点（y_7）预报结果表

k	年份	$\hat{a}^*_{17}(k)$	$\hat{a}^*_{27}(k)$	$\hat{a}^*_{37}(k)$	$\hat{\beta}^*_{17}(k)$	$\hat{\beta}^*_{27}(k)$	$\hat{\beta}^*_{37}(k)$	$\hat{\beta}^*_{47}(k)$	$\hat{y}_7(k)$	$y_7(k)$	δ_7	∇_7 (%)	$\hat{y}_7$ 矩平值	y_7 矩平值	趋势评定
21	1975	0.1363	0.1100	0.0579	0.2626	0.1502	0.1004	0.0568	124	124	0	0	2	2	✓
22	1976	0.1535	0.1265	0.0743	0.2835	0.1675	0.1185	0.0732	141	142	−1	0.7	19	20	✓
23	1977	0.1359	0.1130	0.0553	0.2632	0.1519	0.1008	0.0617	118	120	−2	1.7	−4	−2	✓
24	1978	0.1640	0.1437	0.0853	0.3038	0.1914	0.1371	0.0936	130	160	−30	18.8	8	38	✓
25	1979	0.1775	0.1658	0.1077	0.3032	0.1887	0.1415	0.1187	143	105	38	36.2	21	−17	×
26	1980	0.1245	0.1346	0.0600	0.2793	0.1686	0.1203	0.0693	113	115	−2	1.7	−9	−7	✓
27	1981	0.1196	0.1343	0.0546	0.2709	0.1642	0.1155	0.0646	114	120	−6	5	−8	−2	✓
28	1982	0.1432	0.1568	0.0779	0.2992	0.1882	0.1395	0.0869	129	135	−6	4.4	7	13	✓

表 6.23　西太平洋 8 月份副高西伸脊点（y_8）预报结果表

k	年份	$\hat{a}^*_{18}(k)$	$\hat{a}^*_{28}(k)$	$\hat{a}^*_{38}(k)$	$\hat{\beta}^*_{18}(k)$	$\hat{\beta}^*_{28}(k)$	$\hat{\beta}^*_{38}(k)$	$\hat{\beta}^*_{48}(k)$	$\hat{y}_8(k)$	$y_8(k)$	δ_8	∇_8 (%)	$\hat{y}_8$ 矩平值	y_8 矩平值	趋势评定
21	1975	0.1538	0.1175	0.0665	0.2578	0.1660	0.1259	0.0833	136	135	1	0.7	14	13	✓
22	1976	0.1193	0.0839	0.0383	0.2249	0.1370	0.1024	0.0586	110	110	0	0	−12	−12	✓
23	1977	0.1593	0.1273	0.0818	0.2569	0.1615	0.1327	0.0903	131	142	−11	7.7	9	20	✓
24	1978	0.1495	0.1213	0.0842	0.2560	0.1666	0.1228	0.0841	115	110	5	4.5	−7	−12	✓
25	1979	0.1656	0.1406	0.0995	0.2912	0.2004	0.1517	0.1057	127	130	−3	2.3	5	8	✓
26	1980	0.1574	0.1447	0.0917	0.2696	0.1806	0.1353	0.1030	125	110	15	13.6	3	−12	×
27	1981	0.1027	0.0925	0.0382	0.2223	0.1310	0.0966	0.0682	93	115	−22	19.1	−29	−7	✓
28	1982	0.1675	0.1631	0.1015	0.2831	0.1836	0.1469	0.1286	139	138	1	0.7	17	16	✓

从以上结果可以看出，应用多层递阶预报方法进行夏季各月份副高西伸脊点的自回归长期预报，趋势准确率可达 91.7%，年均相对误差为 6.4%. 从而充分体现了这种方法的优越性.

§6.5 农作物单位面积产量的预报模型

农业生产的情况十分复杂，对作物单位面积产量影响的因素很多，除了一些关系密切的气象因子（可有一定的定量形式表示）外，还有一定条件下的各种社会因素及人为因素（这些往往只是定性的），而产量本身的变化也在一定程度上客观地反映了各种影响因素综合作用的结果.

为了较好地客观地反映错综复杂的实际情况，我们不但应考虑到作物产量自身的发展规律，而且还要适当地考虑与作物生长发育关系密切的气象因子. 例如，对小麦生产影响较大的前期(9，10 月份)降水量及同期(4，5 月份)降水量；与玉米产量有关的同期(6 月份)降水量及 7，8 月份平均温度；与大豆发育关系密切的 7，8 月份降水量及平均气温等等. 对于作物生长系统而言，这些气象因子是不可控制的输入变量. 另外，我们对一般以定性形式表现的社会及人为因素可采用按不同水平分等级评分的方法，请各方面有实际经验的专家们对各项因素进行综合评分，尽可能使之定量化. 我们选取了良种推广、肥料施用、水利排灌、农业机械、其他因素（I 和 II）等六项社会、人为因素作为可以控制（人为可以制约）的输入变量. 而预报对象——农作物的单位面积产量则作为系统的输出变量.

现已掌握这些变量的历史数据见表 6.24～6.26.

表 6.26 中 $u_{11}(k)$, $u_{12}(k)$, $u_{13}(k)$ 分别为第 k 年小麦、玉米、大豆良种推广水平指数，以历年来最高良种推广水平为 100 分，较次的酌情降之；$u_{21}(k)$, $u_{22}(k)$, $u_{23}(k)$ 均为第 k 年该县肥料施用水平指数，以历史上纯施化肥总量及相应的农家肥水平较高、施肥的方法措施较好的 1978 年为 80 分(该年化肥总量 26118 吨)，折合

表6.24 某县小麦、玉米、大豆历年单位面积产量表

时间 k	年份	单位面积产量 y (斤/亩)[1]			时间 k	年份	单位面积产量 y (斤/亩)		
		小麦 y_1	玉米 y_2	大豆 y_3			小麦 y_1	玉米 y_2	大豆 y_3
1	1958	141	269	191	14	1971	331	394	186
2	1959	198	261	212	15	1972	238	273	151
3	1960	119	237	164	16	1973	163	351	193
4	1961	75	139	98	17	1974	248	452	209
5	1962	98	169	119	18	1975	240	396	223
6	1963	99	167	154	19	1976	274	275	122
7	1964	167	222	164	20	1977	251	272	146
8	1965	159	275	209	21	1978	254	470	151
9	1966	215	256	218	22	1979	247	453	183
10	1967	250	294	200	23	1980	264	360	168
11	1968	226	263	175	24	1981	335	447	224
12	1969	223	195	150	25	1982	190	337	102
13	1970	252	321	206					

1) 考虑到资料来源的历史性，在此仍采用了斤、亩等非法定计量单位．1斤＝0.5 kg，1亩＝666.7 m²．

为每326.475吨为1分，按此标准逐年评分；$u_{31}(k)$，$u_{32}(k)$，$u_{33}(k)$均为第k年县水利排灌水平指数，以历史上有效灌溉面积较高的1979年为70分(87830.6亩)，折合1254.7亩为1分，依此逐年评分；$u_{41}(k)$，$u_{42}(k)$，$u_{43}(k)$ 均为第k年县农业机械化水平指数，按各年机耕面积占耕地的百分数判分；$u_{51}(k)$，$u_{52}(k)$，$u_{53}(k)$均为第k年其他因素(I)水平指数，$u_{61}(k)$，$u_{62}(k)$，$u_{63}(k)$ 均为第k年其他因素(II)水平指数，分别按不同年份的不同情况评分。

由此，我们可建立主要粮豆作物单位面积产量预报系统的数学模型如下

$$\mathbf{y}(k) = A(k)\mathbf{y}(k-1,k-2) + B(k)\mathbf{w}(k) + C(k)\mathbf{u}(k) + \mathbf{v}(k) \tag{6.9}$$

其中 $\mathbf{y}(k) = (y_1(k), y_2(k), y_3(k))^\tau$ 为主要粮豆作物单位面积产量(输出)向量；而

表 6.25 某县有关气象因子历年数据表

时间 k	年 份	前期9,10月份降水量(mm) w_{11}	4,5月降水量 w_{21} (mm)	6月降水量 w_{12} (mm)	7,8月降水量 w_{13} (mm)	7,8月温度 w_{22}, w_{23} (℃)
1	1958	102.0	45.2	39.5	212.0	20.6
2	1959	184.3	65.9	106.2	250.7	20.0
3	1960	130.9	86.6	93.7	365.7	20.4
4	1961	80.7	60.1	94.4	445.6	20.5
5	1962	41.8	41.5	63.7	615.3	20.3
6	1963	131.0	125.8	112.2	323.0	20.4
7	1964	132.9	39.5	38.6	271.9	19.9
8	1965	69.8	20.9	18.7	299.2	19.6
9	1966	83.4	45.3	183.8	321.9	20.0
10	1967	80.4	124.4	37.0	206.7	21.1
11	1968	101.4	51.1	42.7	162.5	21.3
12	1969	119.1	88.0	62.3	408.1	19.8
13	1970	110.8	67.5	48.5	212.0	21.1
14	1971	79.9	99.0	73.3	179.9	19.7
15	1972	108.9	47.2	101.7	368.8	19.2
16	1973	139.2	65.4	68.8	352.8	20.7
17	1974	45.0	93.3	125.0	208.9	21.5
18	1975	152.3	68.4	215.9	243.3	20.5
19	1976	49.8	101.9	93.1	149.4	20.0
20	1977	8.9	121.5	130.8	153.4	20.7
21	1978	59.2	53.3	111.1	202.9	20.2
22	1979	87.1	30.6	142.7	171.9	20.2
23	1980	103.0	41.2	113.1	98.9	21.6
24	1981	80.8	74.5	94.9	285.0	20.1
25	1982	54.5	64.8	22.8	277.6	22.6

$$\mathbf{y}(k-1, k-2) = (y_1(k-1), y_1(k-2), y_2(k-1), y_2(k-2), y_3(k-1), y_3(k-2))^\tau$$

$$\mathbf{w}(k) = (w_{11}(k), w_{21}(k), w_{12}(k), w_{22}(k), w_{13}(k), w_{23}(k))^\tau$$

为气象因子输入变量(向量);

$$\mathbf{u}(k) = (u_{11}(k), u_{21}(k), u_{31}(k), u_{41}(k), u_{51}(k), u_{61}(k), u_{12}(k), u_{22}(k), u_{32}(k), u_{42}(k), u_{52}(k), u_{62}(k),$$

表 6.26 某县各有关社会因素水平历年数据表

k	年份	良种推广			肥料	水利	农机	其他因素 I	其他因素 II
		小麦 u_{11}	玉米 u_{12}	大豆 u_{13}	u_{21} u_{22} u_{23}	u_{31} u_{32} u_{33}	u_{41} u_{42} u_{43}	u_{51} u_{52} u_{53}	u_{61} u_{62} u_{63}
1	1958	60	50	58	0.4	43.5	2.8	60	60
2	1959	60	50	58	5	66.2	4.6	50	60
3	1960	60	50	60	0.6	22.7	11.8	50	60
4	1961	60	50	60	1	9.9	12.5	50	60
5	1962	60	50	60	0.2	6.6	5.7	65	60
6	1963	60	60	60	1	6.6	23.6	78	65
7	1964	60	65	70	5	10.6	33.5	78	65
8	1965	70	70	70	4	11.4	32.5	78	65
9	1966	70	70	70	19	9.0	46.7	65	65
10	1967	70	70	70	27	3.8	40.3	60	60
11	1968	75	80	70	28	11.8	25.6	60	60
12	1969	80	80	70	33	14.2	28.5	50	60
13	1970	80	80	70	32	19.9	37.8	50	60
14	1971	80	80	75	28	24.6	44.3	50	60
15	1972	85	80	75	39	25.5	33.2	50	60
16	1973	85	85	80	44	16.1	46.3	50	60
17	1974	90	85	80	32	34.1	44.1	50	60
18	1975	95	90	80	28	30.3	78.2	50	60
19	1976	95	90	80	48	61.5	67.9	50	60
20	1977	95	90	80	29	41.6	66.7	80	75
21	1978	95	92	80	80	42.6	69.6	80	75
22	1979	100	93	85	61	70.0	75.1	80	75
23	1980	100	95	85	79	54.9	78.9	90	80
24	1981	100	100	90	56	34.1	75.7	90	80
25	1982	100	100	90	25	37.8	65.4	95	80

$$u_{13}(k), u_{23}(k), u_{33}(k), u_{43}(k), u_{53}(k), u_{63}(k))^T$$

为社会因素输入向量;

$$\mathbf{v}(k) = (v_1(k), v_2(k), v_3(k))^T$$

为三维的零均值白噪声;

$A(k)$, $B(k)$, $C(k)$ 均为时变参数矩阵,且有

$$A(k)=\begin{pmatrix} \alpha_{11}(k) & \alpha_{21}(k) & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{12}(k) & \alpha_{22}(k) & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_{13}(k) & \alpha_{23}(k) \end{pmatrix}$$

$$B(k)=\begin{pmatrix} \beta_{11}(k) & \beta_{21}(k) & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{12}(k) & \beta_{22}(k) & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{13}(k) & \beta_{23}(k) \end{pmatrix}$$

$$C(k)=$$

$$\left(\begin{matrix} \gamma_{11}(k) & \gamma_{21}(k) & \gamma_{31}(k) & \gamma_{41}(k) & \gamma_{51}(k) & \gamma_{61}(k) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{12}(k) & \gamma_{22}(k) & \gamma_{32}(k) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}\right.$$

$$\left.\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_{42}(k) & \gamma_{52}(k) & \gamma_{62}(k) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_{13}(k) & \gamma_{23}(k) & \gamma_{33}(k) & \gamma_{43}(k) & \gamma_{53}(k) & \gamma_{63}(k) \end{matrix}\right)$$

由多输出系统输出的可分离性，系统(6.9)可分解为如下的单输出系统

$$\begin{aligned} y_j(k) = {} & \alpha_{1j}(k)y_j(k-1) + \alpha_{2j}y_j(k-2) + \beta_{1j}(k)w_{1j}(k) \\ & + \beta_{2j}(k)w_{2j}(k) + \gamma_{1j}(k)u_{1j}(k) + \gamma_{2j}(k)u_{2j}(k) \\ & + \gamma_{3j}(k)u_{3j}(k) + \gamma_{4j}(k)u_{4j}(k) + \gamma_{5j}(k)u_{5j}(k) \\ & + \gamma_{6j}(k)u_{6j}(k) + v_j(k) \qquad j=1,2,3 \end{aligned} \tag{6.10}$$

若置

$$\begin{aligned} \boldsymbol{\varphi}_j^{\tau}(k) = {} & (y_j(k-1),\ y_j(k-2),\ w_{1j}(k),\ w_{2j}(k),\ u_{1j}(k), \\ & u_{2j}(k),\ u_{3j}(k),\ u_{4j}(k),\ u_{5j}(k),\ u_{6j}(k)) \\ \boldsymbol{\theta}_j^{\tau}(k) = {} & (\alpha_{1j}(k),\ \alpha_{2j}(k),\ \beta_{1j}(k),\ \beta_{2j}(k),\ \gamma_{1j}(k),\ \gamma_{2j}(k), \\ & \gamma_{3j}(k),\ \gamma_{4j}(k),\ \gamma_{5j}(k),\ \gamma_{6j}(k)) \end{aligned}$$

则(6.10)式可改写为

$$y_j(k) = \boldsymbol{\varphi}_j^{\tau}(k)\boldsymbol{\theta}_j(k) + v_j(k) \qquad j=1,2,3 \tag{6.11}$$

这就是我们借以进行预报的基本数学模型。

§6.6 试验预报的结果与分析

运用多层递阶预报方法，我们先取 $N=22$；$h=1,2,3$，即利用1979年以前数据，对1980，1981，1982年的单位面积产量进行试验预报。下面以玉米为例说明之。

由历史数据，经参数跟踪，可得到玉米单位面积产量（y_2）的一系列参数估值如表6.27。

对 $\{\hat{\alpha}_{il}(k)\}$，$\{\hat{\beta}_{jl}(k)\}$，$\{\hat{\gamma}_{l2}(k)\}(i=1,2;j=1,2;l=1,2,3,4,5,6)$ 进行分析，可见对 $\{\hat{\alpha}_{12}(k)\}\{\hat{\alpha}_{22}(k)\}\{\hat{\beta}_{12}(k)\}$ 可采用"周期增量法"，其余的采用"多层 AR 模型法"，分别确定它们的预报值。例如，我们可得 $\hat{\alpha}_{12}^*(k)$ 的周期增量估值如下

$$\Delta\hat{\alpha}_{12}^*(k)=\begin{cases}-0.0828 & k=23,28,33,\cdots\\ +0.1084 & k=24,29,34,\cdots\\ -0.0945 & k=25,30,35,\cdots\\ +0.1345 & k=26,31,36,\cdots\\ +0.0356 & k=27,32,37,\cdots\end{cases}$$

又例如，我们可得关于 $\{\hat{\beta}_{22}(k)\}$ 的7阶 AR 预报模型如下

$$\begin{aligned}\hat{\beta}_{22}^*(k)=&0.2023\hat{\beta}_{22}^*(k-1)+0.2035\hat{\beta}_{22}^*(k-2)\\&+0.1548\hat{\beta}_{22}^*(k-3)+0.1531\hat{\beta}_{22}^*(k-4)\\&+0.1038\hat{\beta}_{22}^*(k-5)+0.1042\hat{\beta}_{22}^*(k-6)\\&+0.1036\hat{\beta}_{22}^*(k-7)\end{aligned}$$

…

由参数预报公式得到参数预报估值之后，进一步得出某县玉米1980～1982年单位面积产量预报值及其与真实值比较误差分析结果，见表6.28。

表中 $\delta=\hat{y}_2(k)-y_2(k)$ 为残差，$\nabla=\left|\frac{\delta}{y_2}\right|$ 为相对误差。由以上结果可知，预报的平均相对误差为1.0%。

如果用同样的历史数据，采用一般的多元回归分析方法，则可得到 $y_2(k)$ 的预报公式为

表 6.27 某县玉米单位面积产量（y_2）时变参数估值表

k	$\hat{\alpha}_{12}(k)$	$\hat{\alpha}_{22}(k)$	$\hat{\beta}_{12}(k)$	$\hat{\beta}_{22}(k)$	$\hat{\gamma}_{12}(k)$	$\hat{\gamma}_{22}(k)$	$\hat{\gamma}_{32}(k)$	$\hat{\gamma}_{42}(k)$	$\hat{\gamma}_{52}(k)$	$\hat{\gamma}_{62}(k)$
3	0.3136	0.3171	0.2408	0.2089	0.2218	0.2003	0.2099	0.2051	0.2218	0.2261
4	0.1711	0.1601	0.1840	0.1966	0.1917	0.1997	0.2039	0.1976	0.1917	0.1900
5	0.2566	0.3060	0.2232	0.2090	0.2225	0.1998	0.2080	0.2011	0.2317	0.2270
6	0.2551	0.3047	0.2222	0.2089	0.2219	0.1998	0.2079	0.2009	0.2310	0.2264
7	0.3859	0.4371	0.2524	0.2245	0.2729	0.2037	0.2162	0.2272	0.2921	0.2773
8	0.4722	0.5020	0.2597	0.2321	0.3000	0.2052	0.2207	0.2398	0.3224	0.3025
9	0.2894	0.3544	0.1375	0.2188	0.2535	0.1926	0.2147	0.2087	0.2792	0.2593
10	0.3678	0.4387	0.1489	0.2252	0.2750	0.2009	0.2158	0.2211	0.2976	0.2777
11	0.3002	0.3799	0.1390	0.2203	0.2566	0.1945	0.2131	0.2152	0.2838	0.2639
12	0.1875	0.2538	0.1123	0.2119	0.2223	0.1803	0.2070	0.2030	0.2624	0.2382
13	0.4131	0.5581	0.1685	0.2363	0.3148	0.2173	0.2301	0.2467	0.3202	0.3076
14	0.5181	0.6219	0.1924	0.2427	0.3410	0.2265	0.2381	0.2612	0.3366	0.3273
15	0.1828	0.3487	0.1059	0.2264	0.2729	0.1933	0.2164	0.2330	0.2940	0.2762
16	0.2633	0.4649	0.1262	0.2325	0.2980	0.2063	0.2212	0.2466	0.3088	0.2939
17	0.4583	0.6165	0.1956	0.2444	0.3452	0.2241	0.2401	0.2711	0.3365	0.3272
18	0.2565	0.4598	0.0992	0.2353	0.3050	0.2115	0.2266	0.2362	0.3142	0.3004
19	0.1065	0.2886	0.0639	0.2377	0.2709	0.1934	0.2033	0.2105	0.2953	0.2777
20	0.1248	0.3149	0.0726	0.2291	0.2769	0.1953	0.2060	0.2149	0.3006	0.2827
21	0.4395	0.6331	0.2012	0.2524	0.3834	0.2879	0.2553	0.2954	0.3931	0.3695
22	0.2893	0.5462	0.1556	0.2460	0.3537	0.2684	0.2330	0.2715	0.3676	0.3455

表 6.28　某县玉米单位面积产量试验预报结果及其误差分析表

k	年份	$\hat{\alpha}^*_{12}(k)$	$\hat{\alpha}^*_{22}(k)$	$\hat{\beta}^*_{12}(k)$	$\hat{\beta}^*_{22}(k)$	$\hat{\gamma}^*_{12}(k)$	$\hat{\gamma}^*_{22}(k)$	$\hat{\gamma}^*_{32}(k)$
23	1980	0.1065	0.3306	0.1024	0.2398	0.3213	0.2215	0.2231
24	1981	0.2149	0.4832	0.1427	0.2414	0.3299	0.2257	0.2266
25	1982	0.1204	0.4199	0.1313	0.2435	0.3391	0.2305	0.2314

k	年份	$\hat{\gamma}^*_{42}(k)$	$\hat{\gamma}^*_{52}(k)$	$\hat{\gamma}^*_{62}(k)$	$\hat{y}_2(k)$	$y_2(k)$	δ	∇ (%)
23	1980	0.2474	0.3337	0.3162	355	360	−5	1.4
24	1981	0.2538	0.3388	0.3219	442	447	−5	1.1
25	1982	0.2622	0.3455	0.3288	335	337	−2	0.6

$$
\begin{aligned}
\hat{y}_2(k) = {} & 0.1128\hat{y}_2(k-1) - 0.5785\hat{y}_2(k-2) - 0.1954w_1(k) \\
& + 21.1752w_2(k) - 0.4432u_1(k) + 2.2741u_2(k) \\
& - 0.0608u_3(k) + 3.4096u_4(k) - 1.7957u_5(k) \\
& - 0.8011u_6(k)
\end{aligned}
$$

试验预报结果及其误差分析见表 6.29.

表　6.29

k	年　份	$\hat{y}_2(k)$	$y_2(k)$	δ	∇ (%)
23	1980	392	360	32	8.9
24	1981	303	447	−144	32.2
25	1982	280	337	−57	16.9

从上可知其预报平均相对误差为 19.3%，与多层递阶预报的结果(表 6.29)相比较,显然多层递阶预报的精度要高得多.

类似地,也可得到小麦、大豆 1980～1982 年单位面积产量试验预报结果.

§ 6.7　主要农作物单位面积产量向后 18 年的预报结果

在前节试验预报基础上，可继续运用多层递阶预报方法，取 $N=25$, $h=1, 2, \cdots, 18$. 即用 1982 年以前的数据,对 1983～

表 6.30 各有关气象因子逐年预报值

k	年　份	前期 9,10 月份降水 w_{11} (mm)	4,5 月份降水 w_{21} (mm)	6 月份降水 w_{12} (mm)	7,8 月份降水 w_{13} (mm)	7,8 月份温度 w_{22}, w_{23} (℃)
26	1983	85.0	82.1	131.1	126.1	20.4
27	1984	77.0	85.3	106.5	297.1	20.1
28	1985	90.6	65.0	81.5	366.5	21.4
29	1986	40.8	61.1	89.9	314.8	21.5
30	1987	145.5	102.6	117.4	211.1	21.0
31	1988	86.2	67.3	94.7	167.0	20.6
32	1989	29.3	65.5	59.8	195.8	20.8
33	1990	71.2	59.9	139.3	197.7	20.7
34	1991	94.9	111.2	128.8	160.0	21.4
35	1992	107.0	45.3	117.5	113.3	21.6
36	1993	105.1	105.5	93.3	275.3	20.4
37	1994	90.2	110.2	79.5	212.5	21.6
38	1995	85.7	47.9	158.1	127.4	19.8
39	1996	88.4	78.8	156.4	270.8	19.5
40	1997	111.4	115.5	95.3	314.5	21.2
41	1998	37.1	31.8	146.1	286.4	20.8
42	1999	118.3	78.8	208.6	189.8	20.8
43	2000	72.9	121.4	121.8	145.5	21.2

表 6.31　各社会因素逐年水平预计表

k	年份	良种推广			肥料	水利	农机	其他因素 I	其他因素 II
		小麦 u_{11}	玉米 u_{12}	大豆 u_{13}	u_{21},u_{22},u_{23}	u_{31},u_{32},u_{33}	u_{41},u_{42},u_{43}	u_{51},u_{52},u_{53}	u_{61},u_{62},u_{63}
26	1983	100	100	90	80	65	75	95	85
27	1984	100	100	95	80	65	80	95	90
28	1985	100	100	100	95	80	80	100	95
29	1986	100	100	100	95	80	80	100	95
30	1987	100	100	100	95	80	80	100	95
31	1988	100	100	100	95	80	80	100	95
32	1989	100	100	100	95	80	80	100	95
33	1990	100	100	100	100	90	85	100	100
34	1991	100	100	100	100	90	85	100	100
35	1992	100	100	100	100	90	85	100	100
36	1993	100	100	100	100	90	85	100	100
37	1994	100	100	100	100	90	85	100	100
38	1995	100	100	100	100	95	90	100	100
39	1996	100	100	100	100	95	90	100	100
40	1997	100	100	100	100	95	90	100	100
41	1998	100	100	100	100	95	90	100	100
42	1999	100	100	100	100	95	90	100	100
43	2000	100	100	100	100	100	95	100	100

表 6.32　玉米（y_2）及大豆（y_3）单位面积产量 1983～2000 年预报结果

时间 k	年份	玉米 y_2（斤/亩）	大豆 y_3（斤/亩）	时间 k	年份	玉米 y_2（斤/亩）	大豆 y_3（斤/亩）	时间 k	年份	玉米 y_2（斤/亩）	大豆 y_3（斤/亩）
26	1983	456	195	32	1989	582	274	38	1995	623	288
27	1984	494	214	33	1990	552	232	39	1996	687	299
28	1985	478	288	34	1991	606	270	40	1997	602	356
29	1986	518	230	35	1992	545	244	41	1998	682	295
30	1987	443	267	36	1993	607	315	42	1999	769	334
31	1988	526	238	37	1994	680	252	43	2000	685	303

2000 年的单位面积产量进行预报．为此，我们必须先将预报各年的气象因子的值预报出来，社会因素的各种水平根据实际情况初步确定下来，现将结果列入表 6.30.

表 6.33 小麦单位面积产量（y_1）1983～2000 年预报结果（含预报公式中的时变参数）

时间 k	年份	$\hat{\alpha}^*_{11}(k)$	$\hat{\alpha}^*_{21}(k)$	$\hat{\beta}^*_{11}(k)$	$\hat{\beta}^*_{21}(k)$	$\hat{\gamma}^*_{11}(k)$	$\hat{\gamma}^*_{21}(k)$	$\hat{\gamma}^*_{31}(k)$	$\hat{\gamma}^*_{41}(k)$	$\hat{\gamma}^*_{51}(k)$	$\hat{\gamma}^*_{61}(k)$	$\hat{y}_1(k)$（斤/亩）
26	1983	0.1340	0.3018	0.2082	0.2334	0.2984	0.2030	0.2242	0.2522	0.3133	0.2975	298
27	1984	0.3004	0.4607	0.2082	0.2364	0.3011	0.2038	0.2253	0.2538	0.3154	0.2997	352
28	1985	0.1280	0.2975	0.2091	0.2396	0.3041	0.2052	0.2267	0.2561	0.3178	0.3022	317
29	1986	0.1371	0.3068	0.2093	0.2423	0.3057	0.2066	0.2272	0.2576	0.3193	0.3038	325
30	1987	0.3035	0.4657	0.2083	0.2449	0.3068	0.2051	0.2278	0.2588	0.3204	0.3051	452
31	1988	0.1311	0.3025	0.2070	0.2443	0.3044	0.2061	0.2264	0.2572	0.3179	0.3034	341
32	1989	0.1402	0.3118	0.2062	0.2463	0.3049	0.2060	0.2266	0.2576	0.3176	0.3037	360
33	1990	0.3066	0.4707	0.2062	0.2536	0.3134	0.2056	0.2303	0.2626	0.3258	0.3111	459
34	1991	0.1342	0.3075	0.2061	0.2566	0.3157	0.2062	0.2312	0.2642	0.3277	0.3131	380
35	1992	0.1433	0.3168	0.2059	0.2591	0.3175	0.2067	0.2320	0.2656	0.3293	0.3149	394
36	1993	0.3097	0.4757	0.2054	0.2622	0.3191	0.2069	0.2325	0.2668	0.3307	0.3164	513
37	1994	0.1373	0.3125	0.2047	0.2650	0.3208	0.2069	0 2332	0.2680	0.3320	0.3180	403
38	1995	0.1464	0.3218	0.2041	0.2680	0.3227	0.2069	0.2339	0.2692	0.3335	0.3197	419
39	1996	0.3128	0.4807	0.2037	0.2717	0.3257	0.2069	0.2351	0.2711	0.3361	0.3224	530
40	1997	0.1404	0.3175	0.2034	0.2757	0.3292	0.2071	0.2365	0.2733	0.3392	0.3255	429
41	1998	0.1495	0.3268	0.2031	0.2790	0.3314	0.2073	0.2374	0.2749	0.3410	0.3275	422
42	1999	0.3159	0.4857	0.2027	0.2822	0.3235	0.2075	0.2382	0.2763	0.3428	0.3295	557
43	2000	0.1435	0.3225	0.2022	0·2856	0.3358	0.2076	0.2391	0.2778	0.3447	0.3316	438

类似前面介绍的方法，通过对参数进行跟踪、分析、建立其预报公式，进而由公式(6.13)得到 $y_j(k)(j=1,2,3)$ 的各年预报值，见表 6.32，6.33.

§6.8 小麦单位面积产量的多层递阶自回归预报

由于某种原因，有时我们不能全面地掌握有关预报因子的所需数据，可以不考虑这些因素，而对预报量自身的变化规律进行分析，进而建立起它所满足的 AR 模型或 ARMA 模型.

以下用关于某县的小麦单位面积产量预报的实例，来说明这种途径的应用情况.

我们已得到某县小麦单位面积产量的历史数据，见表 6.34.

表 6.34

年份	1955	1956	1957	1958	1959	1960	1961	1962
k	1	2	3	4	5	6	*	*
$y(k)$ (斤)	162	153	170	243	243	231	197	143

年份	1963	1964	1965	1966	1967	1968	1969	1970
k	7	8	9	10	11	12	13	14
$y(k)$ (斤)	126	162	153	170	243	243	231	197

年份	1971	1972	1973	1974	1975	1976	1977	1978	1979
k	15	16	17	18	19	20	21	*	22
$y(k)$ (斤)	143	126	123	194	187	166	147	31	165

其中 1961，1962 和 1978 年的数据，很明显是反常的，不符合一般规律. 在计算过程中应该将这几个数据剔除.

由上述的历史数据，可确定小麦单位面积产量的动态方程是二阶 AR 模型

$$y_k = \theta_1(k)y_{k-1} + \theta_2(k)y_{k-2} + e_k \tag{6.12}$$

选取时变参数的初值为

$$\hat{\theta}_1(0) = 1.1251 \quad \hat{\theta}_2(0) = -0.0962$$

应用跟踪公式，可得到时变参数的一系列估值见表 6.35。

表 6.35

k	3	4	5	6	7	8
$\hat{\theta}_1(k)$	1.1251	1.0666	1.1728	0.9884	1.0432	1.2964
$\hat{\theta}_2(k)$	−0.0962	−0.1528	−0.0372	−0.2001	−0.1325	0.1552

k	9	10	11	12	13	14
$\hat{\theta}_1(k)$	0.9834	1.0722	1.2513	0.9896	0.9627	0.9166
$\hat{\theta}_2(k)$	−0.0573	0.0367	0.1979	0.0148	−0.0121	−0.0606

k	15	16	17	18	19
$\hat{\theta}_1(k)$	0.8662	0.9283	0.9662	1.[illegible]599	0.8594
$\hat{\theta}_2(k)$	−0.1197	−0.0342	0.0091	0.3098	0.1626

对 $\{\hat{\theta}_1(k)\}$, $\{\hat{\theta}_2(k)\}$ 分别进行分析，由"周期增量法"可得到相应的预报公式如下

$$\hat{\theta}_i^*(k) = \hat{\theta}_i^*(k-1) + \Delta\hat{\theta}_i^*(k) \quad i = 1, 2 \tag{6.13}$$

其中 $\Delta\hat{\theta}_1^*(k)$ 与 $\Delta\hat{\theta}_2^*(k)$ 取值见表 6.36。

表 6.36

k	$\Delta\hat{\theta}_1^*(k)$	$\Delta\hat{\theta}_2^*(k)$
20, 23, 26, 29, 32,…	−0.1032	−0.0731
21, 24, 27, 30, 33,…	+0.0242	+0.0327
22, 25, 28, 31, 34,…	+0.0821	+0.0442

表 6.37

k	年份	$\hat{\theta}_1^*(k)$	$\hat{\theta}_2^*(k)$	$\hat{y}(k)$ (斤/亩)	$y(k)$ (斤/亩)	δ	∇(%)
20	1976	0.7562	0.0895	159	166	−7	4.2
21	1977	0.7804	0.1222	147	147	0	0
22	1979	0.8625	0.1664	153	165	−12	7.3

进而可得小麦单位面积产量的预报公式为

$$\hat{y}(k) = \hat{\theta}_1^*(k)\hat{y}(k-1) + \hat{\theta}_2^*(k)\hat{y}(k-2) \qquad (6.14)$$

应用(6.13),(6.14)式得到的时刻 20, 21, 22 (即 1976, 1977, 1979 年)的时变参数、预报值 $\hat{y}(k)$ 及其误差分析见表 6.37.

由以上结果可知,其预报的平均相对误差为 3.8%.

在此基础上,继续运用公式(6.13)和 (6.14),即可预报出向后 18 年某县小麦单位面积产量的结果,见表 6.38.

表 6.38

k	年 份	$\hat{y}(k)$ (斤/亩)	k	年 份	$\hat{y}(k)$ (斤/亩)
23	1980	152	34	1991	210
24	1981	152	35	1992	199
25	1982	170	36	1993	204
26	1983	158	37	1994	233
27	1984	159	38	1995	222
28	1985	179	39	1996	229
29	1986	168	40	1997	262
30	1987	170	41	1998	252
31	1988	192	42	1999	262
32	1989	181	43	2000	302
33	1990	185			

§ 6.9 关于农业总产值的预报

最后我们给出一个关于某县农业总产值的预报实例. 在这个例子中, 对于时变参数的预报估值的获得我们应用了"周期变化法". 这是因为在相应的预报模型中,时变参数自身的变化是大体上有周期性的缘故.

现已得到某县农业总产值的历史数据,见表 6.39.

其中 $y(k)$ 是时刻 k 的农业总产值, 按 1970 年不变价格计算.

由历史数据, 可以确定某县农业总产值的动态方程为二阶

表 6.39

年份	1955	1956	1957	1958	1959	1960	1961	1962
k	1	2	3	4	5	6	7	8
$y(k)$ (万元)	2031	2321	3056	1902	2602	1592	1662	2115

年份	1963	1964	1965	1966	1967	1968	1969	1970
k	9	10	11	12	13	14	15	16
$y(k)$ (万元)	2549	1866	3069	2771	3671	5010	3331	4692

年份	1971	1972	1973	1974	1975	1976	1977	1978	1979
k	17	18	19	20	21	22	23	24	25
$y(k)$ (万元)	4519	3976	4578	5080	4884	5191	4559	4391	4515

表 6.40

k	3	4	5	6	7	8
$\hat{\theta}_1(k)$	0.6977	0.3092	0.4198	0.2638	0.2644	0.5519
$\hat{\theta}_2(k)$	0.7074	0.4124	0.5901	0.4761	0.4770	0.7524

k	9	10	11	12	13	14
$\hat{\theta}_1(k)$	0.5903	0.2897	0.5084	0.4271	0.4411	0.6458
$\hat{\theta}_2(k)$	0.7825	0.5331	0.8319	0.7825	0.7979	0.9524

k	15	16	17	18	19	20	21
$\hat{\theta}_1(k)$	0.2041	0.2834	0.3822	0.2129	0.3038	0.3940	0.2973
$\hat{\theta}_2(k)$	0.6288	0.7481	0.8182	0.6424	0.7457	0.8240	0.7369

AR 模型(6.12),经选取其时变参数的初值为

$$\hat{\theta}_1(0) = 0.6977 \quad \hat{\theta}_2(0) = 0.7074$$

于是,经参数跟踪,可得到一系列参数估值,见表 6.40.

对 $\{\hat{\theta}_1(k)\}$, $\{\hat{\theta}_2(k)\}$ 进行分析,不难发现,两序列虽然上下摆动,但有一定变化规律,基本上按 8 年一周期变化,故经分段适当整理后,可近似地得到时变参数的预报值,参见表 6.41.

表 6.41

k	$\hat{\theta}_1^*(k)$	$\hat{\theta}_2^*(k)$
22, 30, 38, 46	0.4548	0.6143
23, 31, 39	0.2343	0.6529
24, 32, 40	0.4177	0.6103
25, 33, 41	0.4863	0.6004
26, 34, 42	0.2893	0.7878
27, 35, 43	0.5033	0.4917
28, 36, 44	0.3768	0.6930
29, 37, 45	0.3861	0.7283

表 6.42

k	年 份	$\hat{y}(k)$	$y(k)$	δ	∇ (%)
22	1976	5342	5191	151	2.9
23	1977	4405	4559	−154	3.4
24	1978	5072	4391	681	15.5
25	1979	4873	4515	358	7.9

表 6.43

k	年 份	$\hat{y}(k)$ (万元)	$y^*(k)$ (万元)	k	年 份	$\hat{y}(k)$ (万元)	$y^*(k)$ (万元)
26	1980	5405	5474	37	1991	7205	7297
27	1981	5116	5181	38	1992	7485	7580
28	1982	5673	5745	39	1993	6458	6540
29	1983	5916	5991	40	1994	7266	7358
30	1984	6176	6254	41	1995	7411	7505
31	1985	5310	5377	42	1996	7868	7968
32	1986	5987	6063	43	1997	7604	7701
33	1987	6100	6177	44	1998	8318	8424
34	1988	6481	6563	45	1999	8750	8861
35	1989	6261	6341	46	2000	9089	9750
36	1990	6850	6937				

应用以上参数预报值及 $y(k)$ 的预报分式 (6.14)，可得到1976～1979 年的试验预报结果及其误差，参见表 6.42。

从以上结果可以看出，预报的平均相对误差为 7.4%.

为此，继续应用上述公式，便可预报出本世纪末之前各年的农业总产值. 其结果见表 6.43，其中 $\hat{y}(k)$ 已按 1980 年不变价格换算成 $y^*(k)$.

第七章　工业经济系统的预报

在这一章，我们讨论工业经济系统的预报问题．以油田产量预报、产品销售量预报和工业总产值预报为例，来说明多层递阶等预报方法在工业经济系统中的应用情况．

§7.1　油田产量的预报

1. 预报模型的建立

注水开发油田的动态预报，一直是引起油藏工程界极大兴趣并致力研究解决的问题．动态预报的指标很多，其中尤以油田产油量和产水量最为人们所关注．

油田开发多年以来，产量预报问题一直没有得到令人满意的解决．目前油田规划中所应用的指标预测方法，主要是经验预测方法．这种方法的预报误差较大，而且随着预报时间的增长，误差还将继续增大．其预报结果还受许多主观因素的影响．

随着油田注水开发形势的发展和油田储量参数研究工作的不断深入，对油田规划精度的要求也越来越高．因此，寻求一种基础理论成熟、不受主观因素影响、方法简单、预报精度较高的油田产量预报方法是非常迫切的并具有较大的经济意义．

油田动态预测，是一个多种因素相互制约、相互联系的复杂问题．因此用比较狭隘、孤立的方法来解决这类问题，效果是不会尽如人意的．

分析目前所利用的预测方法，其预报误差较大的重要原因，是把整个(或部分)油田看成了一个非时变的动态系统．而油田事实上是一个时变的动态系统，用一个定常参数的数学模型来描写一个时变参数系统，这就必然存在着较大的差异．而用这种模型进

行预报，其预报误差自然会更大一些.

前面所提出的多层递阶预报方法，其目的就是为了解决时变参数系统的预报问题. 通过大庆等油田的预报实践可以证明，用这种方法进行油田产量预报，能够大大提高预报精度. 此外，这种方法还具有所需信息少、算法简单、便于推广等优点. 它可以在微型机上很容易地实现，完成一次预报只需用机时几分钟. 与数值模拟法、经验统计法比较，不但节省人力，而且节省时间，从而大大提高了工作效率.

对于任一给定的实际问题，用什么样的数学模型进行描述和分析最好，主要取决于问题的物理性质. 注水开发油田的理论和实践都已表明，如果不采取任何措施，一个油田的产油量一定是随着含水上升而递降，而产水量随着含水上升而递增. 对于不同的油田来说，由于地质特征、原油性质及开采方式等差异，递降、递增的特点也不尽相同. 但是只要能准确地预测无措施条件下产油量的递减规律和产水量的递增规律，就会给油田开发规划的制订，带来极大的方便.

然而，油田开发的实际问题很复杂，每年的开井数具有一定的随机性. 每年采取措施的时间和采取措施的工作量，也都具有一定的随机性. 各种措施的有效期问题，也并不十分清楚. 所有这些都为研究油田产油量的自然变化规律带来了一定的困难. 因而将增大我们的预报误差.

分析油田日产油量的变化情况(产量构成曲线)，可以得出如下两点结论:

1. 由于油田追求某一含水条件下的稳产，所以每年必须采取一定的稳产措施. 而稳产措施的采取由于受钻井，实施等诸方面客观条件的约束，又不可能是无限制的.

2. 以年为单位期间，以季(或月)为单位时间步长. 可以看出，若扣除对应时刻(季或月)的措施增产油量，所余的产油量是递减的. 这种包含产油量自然递减和部分措施产量递减的趋势（称为综合递减)也具有一定的规律性.

基于上述两点，从统计观点出发，在以下的讨论中我们假定：各种措施产油量在第二年以及以后逐年的递降，服从“老井”的递降规律。

在上述假设的基础上，以某一年末的产油量（我们确定为1972年）做为产量递减的基点．以年、季或月递减率作为各阶段的递减率．可以算出递减变化的产量时间序列

$$Q_o(1), Q_o(2), \cdots, Q_o(N)$$

N是数据的个数．

以某油田的实际数据为例，其产油量自然递减的历史数据见表 7.1。

表 7.1

k	1	2	3	4	5	6	7	8	9	10
$Q_o(k)$	8.21	8.00	7.84	7.57	7.19	7.24	7.05	6.67	6.35	6.15
k	11	12	13	14	15	16	17	18	19	20
$Q_o(k)$	6.01	5.87	5.64	5.59	5.55	5.56	5.52	5.13	5.03	4.93
k	21	22	23	24	25	26	27	28	29	
$Q_o(k)$	4.91	4.83	4.68	4.60	4.56	4.47	4.36	4.36	4.28	

产水量的历史数据见表 7.2。

表 7.2

k	1	2	3	4	5	6	7	8	9	10
$Q_w(k)$	2.33	2.84	2.81	2.95	3.13	3.53	3.88	4.15	4.65	5.02
k	11	12	13	14	15	16	17	18	19	20
$Q_w(k)$	5.43	6.22	6.79	7.69	8.40	9.31	10.17	10.96	12.33	12.81
k	21	22	23	24	25	26	27	28	29	
$Q_w(k)$	13.67	14.10	14.48	15.26	15.83	17.03	17.31	18.52	19.48	

室内水驱油实验研究结果和油田小井距原油层注水开发全过

程均已表明，随时间的增长产油量以指数趋势递降，产水量以指数趋势递增．结合对前述的数据分析，可以得出自然递减的产油量 $Q_o(k)$ 满足

$$Q_o(k) = \alpha_o(k) + \beta_o(k)k^{\gamma_o(k)} + v_o(k) \tag{7.1}$$

其中 k 是时间，$v_o(k)$ 是随机噪声，$\alpha_o(k)$，$\beta_o(k)$，$\gamma_o(k)$ 是时变参数．

同样可以得出产水量 $Q_w(k)$ 满足

$$Q_w(k) = \alpha_w(k) + \beta_w(k)k^{\gamma_w(k)} + v_w(k) \tag{7.2}$$

其中 $v_w(k)$ 是随机噪声，$\alpha_w(k)$，$\beta_w(k)$，$\gamma_w(k)$ 是时变参数．

我们把模型(7.1)和(7.2)统一的写成

$$Q(k) = \alpha(k) + \beta(k)k^{\gamma(k)} + v(k) \tag{7.3}$$

把第四章所得出的跟踪公式应用于模型(7.3)后有

$$\begin{pmatrix} \hat{\alpha}(k) \\ \hat{\beta}(k) \\ \hat{\gamma}(k) \end{pmatrix} = \begin{pmatrix} \hat{\alpha}(k-1) \\ \hat{\beta}(k-1) \\ \hat{\gamma}(k-1) \end{pmatrix} + \frac{\delta}{a_{k-1}^2 + b_{k-1}^2 + c_{k-1}^2} \begin{pmatrix} a_{k-1} \\ b_{k-1} \\ c_{k-1} \end{pmatrix} \cdot \{Q(k) - \hat{\alpha}(k-1) - \hat{\beta}(k-1)k^{\hat{\gamma}(k-1)}\} \tag{7.4}$$

其中 δ 是适当选取的常数，$\hat{\alpha}(k)$，$\hat{\beta}(k)$，$\hat{\gamma}(k)$ 表示 $\alpha(k)$，$\beta(k)$，$\gamma(k)$ 的估值．而

$$a_{k-1} = 1$$

$$b_{k-1} = k^{\hat{\gamma}(k-1)}$$

$$c_{k-1} = \hat{\beta}(k-1)k^{\hat{\gamma}(k-1)}\ln k$$

按多层递阶预报方法的步骤．当得出关于参数的一系列估值后，应依据这些估值寻求关于参数的"将来"值的预报算法．前已指出，可以根据不同的情况，采取 AR 模型法、定常增量法、周期增量法、广义的 AR 模型法等等．设 $\hat{\alpha}(k+h|k)$，$\hat{\beta}(k+h|k)$，$\hat{\gamma}(k+h|k)$ 是由适当的方法所得到的关于 $\alpha(k+h)$，$\beta(k+h)$，$\gamma(k+h)$ 的预报估值，h 是预报的步长，则关于 $Q(k)$ 的向前 h 步的预报估值 $\hat{Q}(k+h|k)$ 的算法是

$$\hat{Q}(k+h|k) = \hat{\alpha}(k+h|k) + \hat{\beta}(k+h|k)(k+h)^{\hat{\gamma}(k+h|k)} \tag{7.5}$$

2. 实际预报结果

(1) 关于产油量的预报

依据产油量自然递减的历史数据，并应用前面给出的参数跟踪公式(7.4)，可得出时变参数的一系列估值见表 7.3（这里只写出了一部分）.

表 7.3

k	$\hat{\alpha}_o(k)$	$\hat{\beta}_o(k)$	$\hat{\gamma}_o(k)$
15	7.9765	−0.6875	0.4651
16	7.9793	−0.6774	0.4567
17	7.9797	−0.6756	0.4552
18	7.9692	−0.7148	0.4884
19	7.9711	−0.7070	0.4814
20	7.9698	−0.7125	0.4864
21	7.9713	−0.7058	0.4800
22	7.9708	−0.7081	0.4822
23	7.9693	−0.7149	0.4888
24	7.9694	−0.7144	0.4883
25	7.9700	−0.7116	0.4854
26	7.9692	−0.7151	0.4890
27	7.9689	−0.7170	0.4910
28	7.9700	−0.7114	0.4851
29	7.9694	−0.7145	0.4883

从 $\hat{\alpha}_o(k)$, $\hat{\beta}_o(k)$, $\hat{\gamma}_o(k)$ 值的分析可以看出，我们能够应用曲线增量法计算参数的预报估值. 这种预报算法能够写成

$$\begin{aligned}\hat{\alpha}_o(k+h|k) &= \hat{\alpha}_o(k) + h^{\frac{1}{P}}\Delta\alpha_o \\ \hat{\beta}_o(k+h|k) &= \hat{\beta}_o(k) + h^{\frac{1}{Q}}\Delta\beta_o \\ \hat{\gamma}_o(k+h|k) &= \hat{\gamma}_o(k) + h^{\frac{1}{R}}\Delta\gamma_o\end{aligned} \tag{7.6}$$

其中 P, Q, R 是适当取定的常数，$\Delta\alpha_o$, $\Delta\beta_o$, $\Delta\gamma_o$ 是定常增量，它们可由下式确定

$$\Delta\alpha_o = \frac{1}{k - k_o}(\hat{\alpha}_o(k) - \hat{\alpha}_o(k_o))$$

$$\Delta\beta_o = \frac{1}{k - k_o}(\hat{\beta}_o(k) - \hat{\beta}_o(k_o)) \quad k_o < k$$

$$\Delta\gamma_o = \frac{1}{k - k_o}(\hat{\gamma}_o(k) - \hat{\gamma}_o(k_o))$$

k_o 是适当选定的时刻.

这里我们具体地有:

预报起始时间 $k = 29$

$$\hat{\alpha}_o(29) = 7.9694 \quad \hat{\beta}_o(29) = -0.7145$$

$$\hat{\gamma}_o(29) = 0.4883$$

而 $\Delta\alpha_o = 0$, $\Delta\beta_o = -0.0009$, $\Delta\gamma_o = 0.0008$ 从数据 $\{\hat{\alpha}_o(k)\}$. $\{\hat{\beta}_o(k)\}$ 和 $\{\hat{\gamma}_o(k)\}$ 的变化趋势可以看出,应该有 $P = Q = R = 1$,于是可以得出各参数的一系列预报值,从而也就得出了自然递减的产油量预报值,见表 7.4.

表 7.4 预报值、实际值及误差表

预报步长 k	1	2	3	4	5	6	7	8
预报产量 $\hat{Q}_o$	4.1932	4.1188	4.0455	3.9730	3.9012	3.8199	3.7592	3.6889
实际产量 $\hat{Q}_o$	4.2146	4.1288	4.0812	4.0610	3.9336	3.8177	3.7262	3.6951
相对误差(%)	0.507	0.241	0.874	2.1609	0.825	−0.320	−0.882	0.168

从上述结果可以看出,向前 8 步预报的误差相对值最大为 2.16%. 除第 4 步外,其余的相对误差皆小于 1%.

(2) 关于产水量的预报

产水量的预报算法与产油量的完全相同. 这里,我们具体的有

$$k = 29 \qquad \hat{\alpha}_w(29) = 2.3560$$

$$\hat{\beta}_w(29) = 0.3030 \qquad \hat{\gamma}_w(29) = 1.1970$$

关于时变参数的预报,仍然可用曲线增量法. 其增量值分别为

$$\Delta\alpha_w = 0 \quad \Delta\beta_w = 0.0026 \quad \Delta\gamma_w = 0.0022$$

仍然可以取 $P=Q=R=1$，则可得出时变参数的一系列预报值。在此基础上即可得出产水量的预报值，参见表 7.5.

表 7.5

步长 k	1	2	3	4	5	6	7
预报值 $\hat{Q}_w$	20.33	21.28	22.25	23.25	24.28	25.33	26.41
实际值 Q_w	20.66	21.75	22.30	23.43	25.03	26.00	26.97
误 差 (%)	1.11	2.12	0.20	0.77	2.997	2.57	2.08

由上述结果可见，向前 7 步预报，误差相对值最大者是 2.997%。平均相对误差为 1.767%，这个精度已完全满足了工程上的要求。

(3) 注释

在实际工程中，通常希望知道的是实际的产油量，而不是自然递减产量，这就产生了一个预报值的变换问题

另一方面，由于制定油田最优开发规划的需要，要求我们建立产油量和产水量变化的状态方程。如果有了这样的状态方程，那么变换的问题，也就随之解决了。为此，我们置

$$\mathbf{x}(k)=\begin{pmatrix}x_1(k)\\x_2(k)\end{pmatrix}$$

其中 $x_1(k)$ 表示 k 时刻的实际产油量，$x_2(k)$ 表示 k 时刻的实际产水量。$x(k)$ 就是状态变量。于是所需的状态方程可写成

$$\mathbf{x}(k+1)=A(k)\mathbf{x}(k)+B(k)\mathbf{u}(k)$$

其中

$$A(k)=\begin{pmatrix}a_{11}(k) & 0\\ 0 & a_{22}(k)\end{pmatrix}$$

$a_{11}(k)$ 是 k 时刻的产油量的自然递减率，而 $a_{22}(k)$ 是 k 时刻的产水量的自然递增率。它可由下述方法确定

$$Q_w(k+1)=a_{22}(k)Q_w(k)+M(k)$$

其中 $M(k)$ 为采用新措施增加的水量。所以

$$a_{22}(k)=\frac{Q_w(k+1)}{Q_w(k)}-\frac{M(k)}{Q_w(k)}$$

如果 $M(k)$ 已知，则 $a_{22}(k)$ 就完全确定了. 如果 $M(k)$ 未知，我们可以应用 $M(k)/Q_w(k)$ 的估值. 实践表明，在其油田的中含水期，其值在 0.03～0.05 间变化，不会对 $a_{22}(k)$ 有太大影响.

状态方程中的 $\mathbf{u}(k)$ 是在第 $k+1$ 时刻将发挥作用的措施(控制)向量，$B(k)$ 是措施作用矩阵，它是已知的.

如果用 $\hat{\mathbf{u}}(k), \hat{\mathbf{u}}(k+1), \cdots, \hat{\mathbf{u}}(k+h-1)$ 表示相继的 h 个时刻将采取的措施向量. 则我们不难得出实际产量的预报公式

$$\hat{\mathbf{x}}(k+1) = A(k)\mathbf{x}(k) + B(k)\hat{\mathbf{u}}(k)$$

$$\begin{aligned}\hat{\mathbf{x}}(k+2) &= A(k+1)\hat{\mathbf{x}}(k+1) + B(k+1)\hat{\mathbf{u}}(k+1) \\ &= A(k+1)A(k)\mathbf{x}(k) + A(k+1)B(k)\hat{\mathbf{u}}(k) \\ &\quad + B(k+1)\hat{\mathbf{u}}(k+1)\end{aligned}$$

一般地我们有

$$\begin{aligned}\hat{\mathbf{x}}(k+h) = A(k+h-1)\cdots A(k)\mathbf{x}(k) \\ + B(k+h-1)\hat{\mathbf{u}}(k+h-1) \\ + \sum_{i=k}^{k+h-2} A(k+h-1)\cdots A(i+1)B(i)\hat{\mathbf{u}}(i)\end{aligned}$$

如果不采取任何措施，我们有

$$\hat{\mathbf{x}}(k+h) = A(k+h-1)\cdots A(k)\mathbf{x}(k)$$

其中 $A(k)$ 的各元素在将来时刻之值，由一系列的预报值决定.

§7.2 产品销售量的预报

产品市场需求量(销售量)的变化是十分复杂的. 通常一种产品(商品)市场需求量的变化(反映产品的寿命周期)大致可分成以下四个阶段:

1) 试销阶段(产品的发展期)，在这一阶段中，产品刚刚投入市场，还没有被广大购买者所认识. 此时销售量随着市场的开发与扩大，有缓慢增加的趋势.

2) 畅销阶段(产品的成熟期)，一种好的并且是生活或生产经常需要的产品，一经受到广大的消费者(购买者)的欢迎，即进入

了畅销阶段，在这一阶段，市场迅速扩大，销售量将日益增加.

3）饱和阶段（产品的饱和期），由于畅销阶段的刺激，该种产品的产量猛增，使得在市场上出现供求相当或供过于求的形势．这时，该种产品在市场上达到了饱和阶段．从生产厂家来看，后继产品的销路将会受到影响，故不宜再增加该产品的产量．所以，生产厂家要特别注意市场情况的变化.

4）淘汰阶段（产品的衰退期），任何一种消费品的销售量都随社会文明的发展而变化．所以要出现产品不断更新换代现象，如果有物美价廉的新产品代替了老产品，则老产品即到达了淘汰阶段．此时必须随着市场需求量的减少而减少该产品的产量．甚至尽早停止生产.

有些产品的需求情况还要受季节变化、年景变化、风土人情等多种因素的影响.

由上述可见，产品销售量的预报，是一项复杂的工作．在进行预报时，至少应注意以下几个方面的问题：

1）分析该种产品的销售过程，并分清我们所要预报的产品处于市场销售过程的哪一个阶段．从而明了该产品销售量的大概变化趋势.

2）对被预报量进行分析，找出与它有关的主要因素作为预报因子.

3）建立一个适用的预报模型，依据这个模型，用有效的预报方法进行预报.

在预报过程中，尽可能地采用多种预报方法，并对所得结果进行综合分析.

以下我们以自行车销售量的预报为例，来说明多层递阶预报方法在解决销售量预报问题中的应用情况.

由于我们所考虑的地区是以农业人口为主的，即在预报期间内，自行车购买力主要来自农民．所以我们把农民年平均纯收入 $u(k)$ 选为预报因子.

我们得到了有关的历史数据见表7.6.

表 7.6　自行车销售量的历史数据表

k（年）	1	2	3	4	5	6	7
$y(k)$（辆）	356	282	369	234	592	1164	1050
$u(k)$（元）	71.7	86.2	67.4	61.7	71	69.4	57.71
k（年）	8	9	10	11	12	13	14
$y(k)$（辆）	1183	1129	1240	760	960	2310	1960
$u(k)$（元）	71.6	94.87	113.84	99.55	58.49	97.80	107
k（年）	15	16	17	18	19	20	21
$y(k)$（辆）	2460	2859	2177	2824	3423	3221	4436
$u(k)$（元）	79	98.93	94.6	96.6	70.2	84.5	109

通过对表中数据及自行车实际销售情况的分析，可以确定该产品正处于成熟期与饱和期，其销售量的动态方程为

$$y_k = \alpha_1 y_{k-1} + \alpha_2 y_{k-2} + \beta_0 u(k) + \beta_1 u(k-1) + e_k \quad (7.7)$$

其中 y_k 是时刻 k 的自行车销售量，u_k 是农民在时刻 k 的人年平均实际或预计的纯收入，$\alpha_1, \alpha_2, \beta_0, \beta_1$ 皆为时变参数，k 是流动时间，$e(k)$ 是随机噪声。

对于系统(7.7)，可以得出关于它的时变参数的跟踪公式

$$\begin{pmatrix}\hat{\alpha}_1(k)\\ \hat{\alpha}_2(k)\\ \hat{\beta}_0(k)\\ \hat{\beta}_1(k)\end{pmatrix} = \begin{pmatrix}\hat{\alpha}_1(k-1)\\ \hat{\alpha}_2(k-1)\\ \hat{\beta}_0(k-1)\\ \hat{\beta}_1(k-1)\end{pmatrix} + \frac{1}{y_{k-1}^2 + y_{k-2}^2 + u(k)^2 + u(k-1)^2}$$

$$\times \begin{pmatrix}y_{k-1}\\ y_{k-2}\\ u(k)\\ u(k-1)\end{pmatrix}\{y_k - \hat{\alpha}_1(k-1)y_{k-1} - \hat{\alpha}_2(k-1)y_{k-2}$$

$$- \hat{\beta}_0(k-1)u(k) - \hat{\beta}_1(k-1)u(k-2)\} \quad (7.8)$$

其中 $\hat{\alpha}_1(k-1)$，$\hat{\alpha}_2(k-1)$，$\hat{\beta}_0(k-1)$，$\hat{\beta}_1(k-1)$ 表示 $\alpha_1(k-1)$，$\alpha_2(k-1)$，$\beta_0(k-1)$，$\beta_1(k-1)$ 的估值。

适当地选取初值：

$$\hat{\alpha}_1(0) = -0.778 \quad \hat{\alpha}_2(0) = 1.89$$

$$\hat{\beta}_0(0) = 6.02 \qquad \hat{\beta}_1(0) = -5.69$$

应用公式(7.8)得出时变参数的一系列估值见表 7.7。

表 7.7

k	$\hat{\alpha}_1(k)$	$\hat{\alpha}_2(k)$	$\hat{\beta}_0(k)$	$\hat{\beta}_1(k)$
3	−0.7780	1.8903	6.0205	−5.6899
4	−0.7776	1.8905	6.0201	−5.6899
5	−0.7776	1.3905	6.0201	−5.6899
6	0.8838	2.547	6.2160	−5.490
7	−0.1090	2.0423	6.1670	−5.5502
8	−0.620	1.4753	6.132	−5.578
9	−0.5586	1.5302	6.132	−5.578
10	−0.6044	1.4822	6.1325	−5.578
11	−0.6654	1.482	6.132	−5.578
12	−0.6654	1.426	6.127	−5.584
13	−0.703	1.365	6.124	−5.589
14	0.3600	2.2070	6.230	−5.524
15	−0.0509	2.036	6.214	−5.541
16	−0.265	1.699	6.205	−5.558
17	−0.5210	1.4781	6.1960	−5.5670
18	−0.578	1.403	6.1944	−5.569
19	−0.117	1.7635	6.2060	−5.5537

从表 7.7 可以看出，$\hat{\beta}_0(k)$，$\hat{\beta}_1(k)$ 分别收敛到 6.2 和 −5.56 左右 $\hat{\alpha}_1(k)$，$\hat{\alpha}_2(k)$ 在一定范围内变化．可以认为它们是平稳的，对 $\hat{\alpha}_1(k)$，$\hat{\alpha}_2(k)$ 建立 AR 模型．应用 FPE准则．求出 $\alpha_1(k)$ 的模型为 4 阶，即

$$\hat{\alpha}_1(k) = a_1\hat{\alpha}_1(k-1) + a_2\hat{\alpha}_1(k-2) + a_3\hat{\alpha}_1(k-3) + a_4\hat{\alpha}_1(k-4) + e_1(k) \tag{7.9}$$

可以选取未知参数 a_1，a_2，a_3，a_4 的初值如下

$$a_1 = 0.1628 \quad a_2 = 0.2361 \quad a_3 = 0.4444 \quad a_4 = 0.0744$$

从这些初值出发，对 a_1，a_2，a_3，a_4 分别进行递推估算，可求出关于它们的一系列估值．然后取其平均值得

$$a_1 = 0.1141 \quad a_2 = 0.2248 \quad a_3 = 0.4271 \quad a_4 = 0.1457$$

用这样的 a_1, a_2, a_3, a_4 求第 20，21 时刻(即 1977，1978 年) 的 $\hat{\alpha}_1$ 值为

$$\hat{\alpha}_1(20) = -0.4042 \quad \hat{\alpha}_1(21) = -0.3944$$

同理,求出 $\alpha_3(k)$ 的 AR 模型的阶数是 7，并预报出第 20，21 时刻的 α_2 值为

$$\hat{\alpha}_2(20) = 1.7562 \quad \hat{\alpha}_2(21) = 1.7148$$

这样,可得出 20 时刻的自行车销售量的预报值为

$$\hat{y}(20) = (-0.4042) \times 3423 + 1.7562 \times 2824 + 6.2 \times 84.5 - 5.56 \times 70.2 \approx 3704$$

同理求出 21 时刻的自行车销售量的预报值为

$$\hat{y}(21) = 4805$$

平均预报相对误差为 11.5%。

§7.3 工业总产值的预报

1. 关于工业总产值的自回归多层递阶预报

设我们已取得了某县的工业总产值的历史数据,见表 7.8.

表 7.8

年代	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
序号 k	1	2	3	4	5	6	7	8	9	10	11
产值 y_k (万元)	876	1383	1324	1835	2707	2413	2472	2628	2472	3161	4353
年代	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
序号 k	12	13	14	15	16	17	18	19	20	21	22
产值 y_k (万元)	3700	3175	2779	2613	2653	2891	2964	3310	2863	3358	3685
年代	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981
序号 k	23	24	25	26	27	28	29	30	31	32	33
产值 y_k (万元)	4089	4338	4592	4790	5606	5883	6876	6772	7880	8642	9495

依据对这些数据及该县工业经济情况的分析，我们可采用二阶时变参数的 AR 模型来描写该县的工业总产值的动态变化情况。即

$$y_k = \theta_1(k)y_{k-1} + \theta_2(k)y_{k-2} + e_k \tag{7.10}$$

其中 y_k 是时刻 k 的工业总产值，$\theta_1(k)$，$\theta_2(k)$ 是时变参数，e_k 是随机噪声。

$\theta_1(k)$，$\theta_2(k)$ 的估值 $\hat{\theta}_1(k)$，$\hat{\theta}_2(k)$ 由下式确定

$$\begin{pmatrix}\hat{\theta}_1(k)\\ \hat{\theta}_2(k)\end{pmatrix} = \begin{pmatrix}\hat{\theta}_1(k-1)\\ \hat{\theta}_2(k-1)\end{pmatrix} + \frac{1}{y_{k-1}^2 + y_{k-2}^2}\begin{pmatrix}y_{k-1}\\ y_{k-2}\end{pmatrix}\{y_k - \hat{\theta}_1(k-1)y_{k-1} - \hat{\theta}_2(k-1)y_{k-2}\} \tag{7.11}$$

于是我们可以得出二层时间序列 $\boldsymbol{\theta}(k) = (\theta_1(k), \theta_2(k))^\tau$ 的一系列估值

$$\hat{\boldsymbol{\theta}}(1), \hat{\boldsymbol{\theta}}(2), \cdots, \hat{\boldsymbol{\theta}}(N)$$

依据这些数据，我们可建立 $\hat{\theta}_i(k)$ 所满足的 AR 模型

$$\theta_1(k) = a_1\theta_1(k-1) + a_2\theta_1(k-2) + a_3\theta_1(k-3) + a_4\theta_1(k-4) + e_1(k) \tag{7.12}$$

$$\theta_2(k) = b_1\theta_2(k-1) + b_2\theta_2(k-2) - b_3\theta_2(k-3) + b_4\theta_2(k-4) + e_2(k) \tag{7.13}$$

对于(7.12)和(7.13)式中出现的未知参数 $a_1, \cdots, a_4, b_1, \cdots, b_4$ 当然仍可以用类似于(7.11)式的递推算法来确定它们的估值。

在我们的实际问题中，经对所得到的估值列 $\{\hat{a}_i(k)\}$，$\{\hat{b}_i(k)\}$ 进行分析后发现，可以认为它们均是非时变的。

于是按多层递阶方法，从 1977 年以前的数据出发，可得出向后 4 年的试验预报值，参见表 7.9。

表 7.9

年份	真值 $y(k)$	估值 $\hat{y}(k)$	绝对误差 $\Delta = y - \hat{y}$	相对误差 $\lvert\Delta\rvert / y$ (%)
1978	6772	7265	−493	7.2
1979	7880	7959	−79	1.0
1980	8642	8651	−9	0.1
1981	9495	9570	−75	0.8

以上结果表明：

1）对 1978～1981 年的预测结果与真值相比，误差比较小，因此有充分理由认为模型拟合程度是可信的。

2）由于 1977 年产值较 1976 年的产值有大幅度的提高，故预测估值稍偏高。

（1）中期预测

我们以 1949～1981 年的 33 年数据为基础，对 1982～1990 年进行 9 年的中期预测，计算结果见表 7.10。1990 年与 1980 年的产值之比为 2.5。

表 7.10

年份	1980	1983	1985	1990
预测产值(万元)	8642	11232	13541	21238

（2）远期预测

随着年代的推移，任何一种预测方法的误差都会越来越大，这是不言而喻的。为了克服这一不足，减少失真程度，我们采用分段的办法，进行逐级预测。

第一步　我们用中期预测的办法，测得 1982～1990 年的产值。

第二步　将测得的 1982～1990 年数据并入原数据组，再向后 10 年预测。结果见表 7.11。2000 年与 1980 年的产值之比为 5.9。

表 7.11

年份	1991	1995	2000
预测值(万元)	23199	32673	50483

2. 关于工业总产值的常规自适应预报

下述的应用实例说明，常规的自适应预报方法对于一些问题特别是短期预报，也能够获得较好的效果。

我们来考虑某地区的工业总产值的预报问题。把与总产值有密切关系的投资额作为预报因子。设已得到的历史数据见表

7.12.

表 7.12

年　份	1953	1954	1955	1956	1957	1958*	1959*
总产值(亿元)	21.59	23.70	23.53	30.11	34.29	57.41	81.16
投资额(亿元)	4.01	4.93	5.58	5.85	5.33	10.03	11.42
年　份	1960*	1961*	1962	1963	1964	1965	1966
总产值(亿元)	95.74	44.03	40.46	45.92	54.52	71.05	86.72
投资额(亿元)	14.40	5.07	4.16	5.06	6.47	5.88	6.36
年　份	1967*	1968*	1969	1970	1971	1972	1973
总产值(亿元)	78.12	75.03	95.08	115.91	119.53	125.29	130.01
投资额(亿元)	4.36	3.73	5.42	6.16	5.83	6.62	8.11

年　份	1974	1975	1976*	1977	1978
总产值(亿元)	144.59	160.69	169.52	184.20	199.44
投资额(亿元)	8.81	10.97	8.26	8.89	13.25

用 y_k 表示第 k 年的工业总产值，$u(k)$ 表示第 k 年的工业总投资额，近似的取 $n=1$，$m=2$，则相应的 CAR 模型是

$$y_{k+1}+a_0y_k+a_1y_{k-1}$$
$$=b_0u(k)+b_1u(k-1)+b_2u(k-2)+e(k+1)$$

我们采用如下的自适应预报算法

$$\hat{y}(k+1)=\boldsymbol{\phi}(k)^\tau\boldsymbol{\theta}^\tau(k-1)$$
$$\hat{\boldsymbol{\theta}}(k)=\hat{\boldsymbol{\theta}}(k-1)+M(k)\{y_{k+1}-\boldsymbol{\phi}(k)^\tau\hat{\boldsymbol{\theta}}(k-1)\}$$
$$M(k)=\frac{P(k-1)\boldsymbol{\phi}(k)}{\alpha+\boldsymbol{\phi}(k)^\tau P(k-1)\boldsymbol{\phi}(k)}$$
$$P(k)=\frac{1}{\alpha}[I-M(K)\boldsymbol{\phi}(K)^\tau]P(k-1)$$

为应用上述的预报公式，首先根据部分统计资料，用最小二乘法确定 $\hat{\boldsymbol{\theta}}(k)$ 和 $P(k)$ 的初始估值 $\hat{\boldsymbol{\theta}}(k_0)$，$P(k_0)$，在计算中，应该把明显反常的数据剔除出去． 例如，在前面的统计资料中，1958，1959，1960 和 1961 年的数据，显然不符合一般的规律． 而 1967

和 1968 年两年的资料也是反常的. 在计算过程中,应该把这些数据剔除出去.

应用上述的自适应预报算法，依据前述的资料，对 1972～1978 年的工业总产值进行逐年的预报,其结果见表 7.13.

表 7.13

年　份	1972	1973	1974	1975	1976	1977	1978
预报值(亿元)	130.07	139.41	148.49	160.07	181.61	181.09	205.85
真实值(亿元)	125.29	135.01	144.59	160.69	169.52	184.20	199.44

从上述的计算结果与简单分析的初步结论可以看出，我们所提出的预报方法,是有实用价值的.

此外,由数理统计的有关理论得知,我们的工业总产值的真实值 y_k 与预报值 $\hat{y}(k)$ 之间应该近似地满足下述关系

$$P\{\hat{y}(k)-2\hat{\sigma}<y_k<\hat{y}(k)+2\hat{\sigma}\}=0.95$$

这就是所谓 2σ 原则,其中 $\hat{\sigma}^2$ 是 $\mathrm{Var}\{e(k)\}$ 的估值，它可以通过多组数据所算得的残差来近似估计,我们得出

$$\hat{\sigma}=2.5$$

由此不难看出,除 1976 年的极特殊情况外,在信度 $\alpha=0.05$,即置信水平 $1-\alpha=0.95$ 时，真实值基本上都落实在我们的预报区间

$$[\hat{y}-2\hat{\sigma},\hat{y}+2\hat{\sigma}]$$

当然,如果预报的步长加大,其效果就不会如此令人满意.

第八章　多层递阶方法在自适应控制中的应用

由于本章所介绍的每种方法都涉及到模型参数的自适应预报问题，对此我们采用了多层递阶方法，所以有时也称这些自适应控制方法为多层递阶参数自适应控制.

§8.1　参数预报自适应控制及其在农业中的应用

我们考虑如下系统的自适应控制问题

$$\begin{aligned} & y_k + \alpha_1(k)y_{k-1} + \cdots + \alpha_n(k)y_{k-n} \\ & = \beta_0(k)u(k-h) + \beta_1(k)u(k-h-1) + \cdots \\ & \quad + \beta_m(k)u(k-h-m) + e_k \end{aligned} \tag{8.1}$$

其中 e_k 是零均值的白噪声，h 是已知的正整数，$\alpha_1(k), \cdots, \alpha_n(k)$，$\beta_0(k), \beta_1(k), \cdots, \beta_m(k)$ 是未知的时变参数．置

$$\begin{aligned} \boldsymbol{\phi}(k)^\tau = [& -y_{k-1}, \cdots, -y_{k-n}, u(k-h), \cdots, \\ & u(k-h-m)] \end{aligned}$$

$$\boldsymbol{\theta}(k)^\tau = [\alpha_1(k), \cdots, \alpha_n(k), \beta_0(k), \beta_1(k), \cdots, \beta_m(k)]$$

则公式(8.1)可以写成

$$y_k = \boldsymbol{\phi}(k)^\tau \boldsymbol{\theta}(k) + e_k$$

设我们已得到了时变参数 $\boldsymbol{\theta}(k)$ 的一系列估值 $\hat{\boldsymbol{\theta}}(1)$，$\hat{\boldsymbol{\theta}}(2), \cdots, \hat{\boldsymbol{\theta}}(N)$，此处$N$是“现在时刻”．从这些数据出发，应用多层递阶预报方法，可以得出参数 $\boldsymbol{\theta}(k)$ 的一系列预报值

$$\begin{aligned} \hat{\boldsymbol{\theta}}(N+d|N)^\tau = [& \hat{\alpha}_1(N+d|N), \cdots, \hat{\alpha}_n(N+d|N), \\ & \hat{\beta}_0(N+d|N)\hat{\beta}_1(N+d|N), \cdots, \\ & \hat{\beta}_m(N+d|N)] \end{aligned}$$

$$d = 1, 2, \cdots, h$$

应用这些预报值和观测数据

$$\{y_0, y_1, \cdots, y_N\} \ \{u(0), u(1), \cdots, u(N)\}$$

可以得出 y_k 的一系列预报值 $\hat{y}(N+d), d=1, 2, \cdots, h$

$$\begin{aligned}
\hat{y}(N+1) = & -\hat{\alpha}_1(N+1|N)y_N - \cdots - \hat{\alpha}_n(N+1|N)y_{N+1-n} \\
& + \hat{\beta}_0(N+1|N)u(N+1-h) + \cdots \\
& + \hat{\beta}_m(N+1|N)u(N+1-h-m) \\
\hat{y}(N+2) = & -\hat{\alpha}_1(N+2|N)\hat{y}(N+1) \\
& - \hat{\alpha}_2(N+2|N)y_N - \cdots \\
& - \hat{\alpha}_n(N+2|N)\hat{y}(N+2-n) \\
& + \hat{\beta}_0(N+2|N)u(N+2-h) + \cdots \\
& + \hat{\beta}_m(N+2|N)u(N+2-h-m)
\end{aligned}$$

$\cdots\cdots$

$$\begin{aligned}
\hat{y}(N+h) = & -\hat{\alpha}_1(N+h|N)y^*(N+h-1) - \cdots \\
& - \hat{\alpha}_n(N+h|N)y^*(N+h-n) \\
& + \hat{\beta}_0(N+h|N)u(N) \\
& + \hat{\beta}_1(N+h|N)u(N-1) + \cdots \\
& + \hat{\beta}_m(N+h|N)u(N-m)
\end{aligned} \tag{8.2}$$

其中

$$y^*(N+h-i) = \begin{cases} y_{N+h-i} & i \geq h \\ \hat{y}(N+h-i) & i < h \end{cases}$$

$$i = 1, 2, \cdots, n$$

设 k 是现在时刻，并已有数据

$$\{y_0, \cdots, y_k\} \quad \{u(0), \cdots, u(k-1)\}$$

如果 r 是系统在时刻 $k+h$ 希望的输出值，则由 (8.2) 式不难得出，我们所寻求的参数预报自适应控制律是

$$\begin{aligned}
r = & -\hat{\alpha}_1(k+h|k)y^*(k+h-1) - \cdots \\
& - \hat{\alpha}_n(k+h|k)y^*(k+h-n) + \hat{\beta}_0(k+h|k)u(k) \\
& + \hat{\beta}_1(k+h|k)u(k-1) + \cdots \\
& + \hat{\beta}_m(k+h|k)u(k-m)
\end{aligned}$$

式中除 $u(k)$ 外，其他量皆是已知的．上式又可写成明显的**形式**

$$u(k)=\frac{1}{\hat{\beta}_0(k+h|k)}\{r+\hat{\alpha}_1(k+h|k)y^*(k+h-1)+\cdots$$
$$+\hat{\alpha}_n(k+h|k)y^*(k+h-n)$$
$$-\hat{\beta}_1(k+h|k)u(k-1)-\cdots$$
$$-\hat{\beta}_m(k+h|k)u(k-m)\}$$

此处假定 $\hat{\beta}_0(k+h|k)\neq 0$. 为保证这一点在参数估计和预报过程中可取 $\hat{\beta}_0(k+h|k)=\beta_0$(常数).

以下给出的是参数预报自适应控制方法在农业上应用的实例.

设已知小麦单位面积产量的预报模型是

$$y_k=\alpha_1(k)y_{k-1}+\alpha_2(k)y_{k-2}+\beta_1(k)w_1(k)+\beta_2(k)w_2(k)$$
$$+\gamma_1(k)u_1(k)+\gamma_2(k)u_2(k)+\gamma_3(k)u_3(k)$$
$$+\gamma_4(k)u_4(k)+\gamma_5(k)u_5(k)+\gamma_6(k)u_6(k)+e_k \quad (8.3)$$

其中 y_k 是小麦单位面积产量, $w_1(k)$ 与 $w_2(k)$ 为与之有关的气象因子, $u_1(k),\cdots,u_6(k)$ 分别表示良种推广、肥料施用量、水利化程度、机械化程度、农村经济政策、生产组织形式等因素,其中的非数量因素已经过"评分"方法进行了数量化. e_k 是零均值的不相关噪声. $\alpha_1(k),\cdots,\gamma_6(k)$ 等为时变参数. 依据历史数据,由模型(8.3)出发应用多层递阶方法可求出诸时变参数的向前一步的预报值

$$\hat{\alpha}_1(k|k-1)=0.1280 \quad \hat{\alpha}_2(k|k-1)=0.2975$$
$$\hat{\beta}_1(k|k-1)=0.2091 \quad \hat{\beta}_2(k|k-1)=0.2396$$
$$\hat{\gamma}_1(k|k-1)=0.3041 \quad \hat{\gamma}_2(k|k-1)=0.2052$$
$$\hat{\gamma}_3(k|k-1)=0.2267 \quad \hat{\gamma}_4(k|k-1)=0.2561$$
$$\hat{\gamma}_5(k|k-1)=0.3178 \quad \hat{\gamma}_6(k|k-1)=0.3022$$

两个气象因子的多层递阶预报值分别为

$$\hat{w}_1(k|k-1)=90.6 \quad \hat{w}_2(k|k-1)=65$$

假定已知: 良种推广水平 $u_1(k)=100$, 水利化程度 $u_3(k)=85$ (%), 机械化程度 $u_4(k)=85$(%),农村经济政策和生产组织形式分别被"评分",其值为: $u_5(k)=100$, $u_6(k)=95$, 问题是肥料

用量达到怎样的程度才能使小麦单位面积产量达到 400 斤. 历史数据给出前两年小麦单位面积产量分别为 352 斤和 298 斤.

应用参数预报自适应控制算法不难得出:

$$u_2(k) = 487 \text{ 斤}$$

这个结果与历史数据比较后可见, 为了使小麦单产达到 400 斤,必须大大地增加施肥量.

§8.2 稳定的参数自适应控制

考虑下列形式模型的系统的自适应控制问题:

$$A(q^{-1})y_k = B(q^{-1})u(k) + C_1^*(q^{-1})s_1(k) + \cdots + C_p^*(q^{-1})s_p(k) + v_k \qquad (8.4)$$

其中 y_k 是输出, $u(k)$ 是输入, $s_1(k), \cdots, s_p(k)$ 是可观测的干扰, v_k 是模型的非明确随机部分, q^{-1} 是一步延迟算子,而

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$

$$B(q^{-1}) = b_1q^{-1} + \cdots + b_mq^{-m}$$

$$C_i^*(q^{-1}) = c_1^*(i)q^{-1} + \cdots + c_{h_i}^*(i)q^{-h_i} \quad i = 1, \cdots, p$$

参数 $c_1^*(i), \cdots, c_{h_i}^*(i)$ 可能是时变的. 只要置

$$y^*(k) = A(q^{-1})y_k - B(q^{-1})u(k)$$

$$\boldsymbol{\phi}_s(k)^\tau = [s_1(k-1), \cdots, s_1(k-h_1), s(k-1)\cdots s_2(k-h_2)\cdots s_p(k-1)\cdots s_p(k-h_p)]$$

$$\boldsymbol{\eta}(k) = [c_1^*(1)\cdots c_{h_1}^*(1)\cdots c_1^*(p)\cdots c_{h_p}^*(p)]$$

则模型(8.4)可写成

$$y^*(k) = \boldsymbol{\phi}_s(k)^\tau\boldsymbol{\eta}(k) + v_k$$

根据前节的结论,必有随机时变参数 $\theta(k)$ 使得

$$y^*(k) = \boldsymbol{\phi}_s(k)^\tau\boldsymbol{\theta}(k) \qquad (8.5)$$

其中 $\boldsymbol{\theta}(k) = [c_1(1)\cdots c_{h_1}(1)\cdots c_1(p)\cdots c_{h_p}(p)]$ 故在此只须考虑如下形式的系统

$$A(q^{-1})y_k = B(q^{-1})u(k) + C_1(q^{-1})s_1(k)\cdots + C_p(q^{-1})s_p(k) \qquad (8.6)$$

此处

$$C_i(q^{-1}) = c_1(i)q^{-1} + \cdots + c_{h_i}(i)q^{-h_i} \quad i = 1, 2, \cdots, p$$

$c_1(i), \cdots, c_{h_i}(i)$ $(i = 1, 2, \cdots, p)$ 是时变随机参数，置

$$\begin{aligned}\boldsymbol{\phi}(k)^\tau = [&-y_{k-1}, \cdots, -y_{k-n}, u(k-1), \cdots, u(k-m),\\ &s_1(k-1), \cdots, s_1(k-h_1), \cdots, s_p(k-1), \cdots,\\ &s_p(k-h_p)]\end{aligned}$$

$$\begin{aligned}\boldsymbol{\xi}(k)^\tau = [&a_1, \cdots, a_n, b_1, \cdots, b_m, c_1(1), \cdots, c_{h_1}(1), \cdots,\\ &c_1(p), \cdots, c_{hp}(p)]\end{aligned}$$

则模型(8.6)可以写成

$$y_k = \boldsymbol{\phi}(k)^\tau \boldsymbol{\xi}(k)$$

从上述模型出发我们提出稳定的参数自适应控制算法如下：

1）寻求保证系统稳定的系统参数的初始估值.

任取一组使系统稳定的参数的初始值 $\hat{\xi}(0)$，应用估值公式

$$\hat{\boldsymbol{\xi}}(k) = \hat{\boldsymbol{\xi}}(k-1) + \frac{1}{\|\boldsymbol{\phi}(k)\|^2}\boldsymbol{\phi}(k)\{y_k - \boldsymbol{\phi}(k)^\tau \boldsymbol{\xi}(k-1)\}$$

依据离线的观测数据，对参数 $\boldsymbol{\xi}(k)$ 进行离线的递推估算. 如果有某组参数估值已超出了系统的稳定域，递推估算即行终止，并用以前各组参数估值的平均值 $\hat{\xi}(k_0)$ 作为自适应算法的参数初始值.

2）把所求得的参数估值的初值分成两部分

$$\hat{\boldsymbol{\xi}}_1(k_0)^\tau = [\hat{a}_1(k_0), \cdots, \hat{a}_n(k_0), \hat{b}_1(k_0), \cdots, \hat{b}_m(k_0)]$$

$$\begin{aligned}\hat{\boldsymbol{\xi}}_2(k)^\tau = [&\hat{c}_1(1, k_0), \cdots, \hat{c}_{h_1}(1, k_0), \cdots, \hat{c}_1(p, k_0), \cdots,\\ &\hat{c}_{h_p}(p, k_0)]\end{aligned}$$

置

$$A_{k_0}(q^{-1}) = 1 + \hat{a}_1(k_0)q^{-1} + \cdots + \hat{a}_n(k_0)q^{-n}$$

$$B_{k_0}(q^{-1}) = \hat{b}_1(k_0)q^{-1} + \cdots + \hat{b}_m(k_0)q^{-m}$$

显然多项式 $A_{k_0}(q^{-1})$，$B_{k_0}(q^{-1})$ 都是稳定的. 令

$$y_0^*(k) = A_{k_0}(q^{-1})y_k - B_{k_0}(q^{-1})u(k)$$

则自适应控制系统的模型是

$$y_0^*(k) = \boldsymbol{\phi}_s(k)^\tau \xi_2(k) = \boldsymbol{\phi}_s(k)^\tau \theta(k) \tag{8.7}$$

此处

$$\theta(k)^{\tau}=\xi_2(k)^{\tau}=[c_1(1,k),\cdots,c_{h_1}(1,k),\cdots,$$
$$c_1(p,k),\cdots,c_{h_p}(p,k)]$$
$$\phi_s(k)^{\tau}=[s_1(k-1)\cdots s_1(k-h_1)\cdots s_p(k-1)\cdots$$
$$s_p(k-h_p)]$$

3) 应用估值公式

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{1}{\|\phi_s(k)\|^2}\phi_s(k)\{y_0^*(k)-\phi_s(k)^{\tau}\hat{\theta}(k-1)\} \tag{8.8}$$

在线地对时变随机参数 $\theta(k)$ 进行估值，并应用多层递阶预报算法，对 $\hat{\theta}(k)$ 进行向前一步预报，用 $\hat{\theta}(k+1|k)$ 表示所得的预报值.

4) 依据模型(8.7)可导出稳定的参数自适应控制律如下

$$y_0+\hat{a}_1(k_0)y_{k-1}+\cdots+\hat{a}_n(k_0)y_{k-n}=\hat{b}_1(k_0)u(k-1)+\cdots$$
$$+\hat{b}_m(k_0)u(k-m)$$
$$+\phi_s(k)^{\tau}\hat{\theta}(k|k-1)$$

其中 y_0 是给定的希望值，或者是

$$u(k-1)=\frac{1}{\hat{b}_1(k_0)}\{y_0+\hat{a}_1(k_0)y_{k-1}+\cdots+\hat{a}_n(k_0)y_{k-n}$$
$$-\hat{b}_2(k_0)u(k-2)-\cdots-\hat{b}_m(k_0)u(k-m)$$
$$-\phi_s(k)^{\tau}\hat{\theta}(k|k-1)\} \tag{8.9}$$

以下我们对这种控制算法在应用中可能达到的控制精度进行分析：

如果用 $\theta^*(k)$ 表示系统的参数 $\theta(k)$ 的真值，$u^*(k-1)$ 表示在当前的条件下，为使系统输出 $y(k)$ 达到希望值 y_0 时所需采取的真实控制量，则我们应该有

$$y_0+\hat{a}_1(k_0)y_{k-1}+\cdots+\hat{a}_n(k_0)y_{k-n}=\hat{b}_1(k_0)u^*(k-1)$$
$$+\hat{b}_2(k_0)u(k-2)+\cdots$$
$$+\hat{b}_m(k_0)u(k-m)$$
$$+\phi_s(k)^{\tau}\theta^*(k)$$

如果置

$$\begin{aligned} y_0^*(k) = y_0 + \hat{a}_1(k_0)y_{k-1} + \cdots + \hat{a}_n(k_0)y_{k-n} \\ - \hat{b}_1(k_0)u^*(k-1) - \hat{b}_2(k_0)u(k-2) - \cdots \\ - \hat{b}_m(k_0)u(k-m) \end{aligned}$$

则有 $y_0^*(k) = \boldsymbol{\phi}_s(k)^\tau \boldsymbol{\theta}^*(k)$

另一方面,由 $\hat{\boldsymbol{\theta}}(k-1)$ 和上述模型，并应用递推公式 (8.8) 可求得 $\hat{\boldsymbol{\theta}}(k)$, 不难证明它也满足

$$y_0^*(k) = \boldsymbol{\phi}_s(k)^\tau \hat{\boldsymbol{\theta}}(k) \tag{8.10}$$

进一步用 $\hat{u}(k-1)$ 表示由控制策略(8.9) 算得的 $u(k-1)$ 在系统的执行机构上实现的值，与之对应的系统在 k 时刻的真实输出表示为 $\hat{y}_0$, 于是我们有

$$\begin{aligned} \hat{y}_0 + \hat{a}_1(k_0)y_{k-1} + \cdots + \hat{a}_n(k_0)y_{k-n} = \hat{b}_1(k_0)\hat{u}(k-1) \\ + \hat{b}_2(k_0)u(k-2) + \cdots \\ + \hat{b}_m(k_0)u(k-m) \\ + \boldsymbol{\phi}_s(k)^\tau \hat{\boldsymbol{\theta}}(k|k-1) \end{aligned}$$

由(8.10)式中减去此式,可得

$$y_0^* - \hat{y}_0 = \boldsymbol{\phi}_s(k)^\tau \Delta\boldsymbol{\theta}(k) + \hat{b}_1(k_0)\Delta u(k-1)$$

其中 $\Delta\boldsymbol{\theta}(k) = \hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}(k|k-1)$

$$\Delta u(k-1) = u^*(k-1) - \hat{u}(k-1)$$

至此,我们证明了下述定理:

定理 8.1 应用稳定的参数自适应控制律于真实系统,所得出的真实误差是

$$y_0^* - \hat{y}_0 = \boldsymbol{\phi}_s(k)^\tau \Delta\boldsymbol{\theta}(k) + \hat{b}_1(k_0)\Delta u(k-1)$$

这个定理说明，我们所提出的自适应控制律的控制精度依赖于时变参数估值的预报算法的精度和实际系统执行机构的精度.

多项式 $A_{k_0}(q^{-1})$ 和 $B_{k_0}(q^{-1})$ 的稳定性在一定条件下保证了自适应控制系统的稳定性.

§8.3 含未知部分的自适应控制模型的补偿方法

这里，我们假定所考虑的动态系统的模型含有未知部分，它可以写成如下的形式

$$A(q^{-1})y_k = q^{-h}B(q^{-1})u(k) + U(k) + C(q^{-1})e_k \quad (8.11)$$

其中 y_k 是一维输出，$u(k)$ 是输入，不失一般性，也可以假定它是一维的，$U(k)$ 是模型的未知部分，但其中不包含可控变量，e_k 是零均值的白噪声.

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$
$$B(q^{-1}) = b_0 + b_1q^{-1} + \cdots + b_mq^{-m}$$
$$C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_rq^{-r}$$

q^{-1} 是一步延迟算子，$a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, \cdots, c_r$ 是模型的未知参数，$h \geqslant 0$.

如果置

$$\boldsymbol{\eta}(k) = [a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, \cdots, c_r, U(k)]^\tau$$
$$\boldsymbol{\varphi}(k) = [-y_{k-1}, \cdots, -y_{k-n}, u(k-h), \cdots, u(k-m-h)$$
$$e_{k-1}, \cdots, e_{k-r}, 1]^\tau$$

则模型(8.11)能够写成

$$y_k = \boldsymbol{\varphi}(k)^\tau \boldsymbol{\eta}(k) + e_k$$

未知参数 $\boldsymbol{\eta}(k)$ 可以用下述的递推算法进行估计

$$\hat{\boldsymbol{\eta}}(k) = \hat{\boldsymbol{\eta}}(k-1) + M(k)\{y_k - \hat{\boldsymbol{\varphi}}(k)^\tau \hat{\boldsymbol{\eta}}(k-1)\} \quad (8.12a)$$

$$M(k) = \frac{P(k-1)\hat{\boldsymbol{\varphi}}(k)}{\lambda_k + \hat{\boldsymbol{\varphi}}(k)^\tau P(k-1)\hat{\boldsymbol{\varphi}}(k)}$$

$$P(k) = \frac{1}{\lambda_k}[I - M(k)\hat{\boldsymbol{\varphi}}(k)^\tau]P(k-1)$$

$$\hat{\boldsymbol{\varphi}}(k) = [-y_{k-1}, \cdots, -y_{k-n}, u(k-h), \cdots,$$
$$u(k-m-h), \varepsilon(k-1), \varepsilon(k-2), \cdots,$$
$$\varepsilon(k-r), 1]^\tau$$

$\varepsilon(k) = y_k - \hat{\boldsymbol{\varphi}}(k)^\tau \hat{\boldsymbol{\eta}}(k)$， λ_k 是遗忘因子

由 $\eta(k)$ 的估值 $\hat{\eta}(1), \cdots, \hat{\eta}(k)$ 可以得出 $U(k)$ 的估值 $\hat{U}(1), \hat{U}(2), \cdots, \hat{U}(k_f)$.

这里 k_f 是现在时刻，置

$$U = \frac{1}{k_f}\sum_{k=1}^{k_f}\hat{U}(k)$$

以及

$$d = \begin{cases} U & U \neq 0 \\ \text{适当的非零值} & U = 0 \end{cases}$$

再置

$$\hat{a}(k) = \frac{\hat{U}(k)}{d}$$

则我们的模型可以写成

$$A(q^{-1})y_k = q^{-h}B(q^{-1})u(k) + \alpha(k)d + C(q^{-1})e_k \quad (8.12)$$

令

$$\eta_1(k) = [a_1, \cdots, a_n, b_0, b_1, \cdots, b_m, c_1, \cdots, c_r]^\tau$$

$G = \{\eta_1(k)$；以 $\eta_1(k)$ 的相应分量为系数的多项式 $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$ 皆是稳定的$\}$

即 G 是参数的稳定域，再置

$$\eta_*(k) = [\eta_1(k)^\tau, \alpha(k)]^\tau$$

$$\hat{\varphi}_*(k) = [-y_{k-1}, \cdots, -y_{k-n}, u(k-h), \cdots, u(k-m-h)\varepsilon(k-1), \cdots, \varepsilon(k-r), d]^\tau$$

我们选取初值 $\hat{\eta}_*(0) = [\hat{\eta}_1(0)^\tau, \hat{a}(0)]$，使得 $\hat{\eta}_1(0) \in G$ 应用算法

$$\eta_*(k) = \hat{\eta}_*(k-1) + M(k)\{y_k - \hat{\varphi}_*(k)^\tau\hat{\eta}_*(k-1)\}$$

$$M(k) = \frac{P(k-1)\varphi_*(k)}{\lambda_k + \varphi_*(k)^\tau P(k-1)\varphi_*(k)}$$

$$P(k) = \frac{1}{\lambda_k}[I - M(k)\varphi_*(k)^\tau]P(k-1)$$

对未知参数 $\eta_*(k)$ 进行估计．但要求 $\hat{\eta}_1(k) \in G$ 如果有 N_0，使 $\hat{\eta}_1(N_0+1) \notin G$，而 $\hat{\eta}_1(N_0) \in G$．于是我们固定 $A(q^{-1})$, $B(q^{-1})$,

$C(q^{-1})$ 的系数为

$$a_1=\hat{a}_1(N_0),\ \cdots,\ a_n=\hat{a}_n(N_0)$$
$$b_0=\hat{b}_0(N_0),\ \cdots,\ b_m=\hat{b}_m(N_0)$$
$$c_1=\hat{c}_1(N_0),\ \cdots,\ c_r=\hat{c}_r(N_0)$$

于是在以下的讨论中,我们恒假定在模型

$$A(q^{-1})y_k=q^{-h}B(q^{-1})u(k)+\alpha(k)d+C(q^{-1})e_k$$

中,$A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$ 皆为已知的稳定的多项式. d 是已知的常数, 仅有 $\alpha(k)$ 是未知的参数, $\alpha(k)d$ 是模型未知部分的补偿项.

§8.4 补偿自适应最小方差控制

这里我们所考虑的系统是

$$A(q^{-1})y_k=q^{-h}B(q^{-1})u(k)+\alpha(k)\overline{N}+C(q^{-1})e_k \quad (8.13)$$

其中 $A(q^{-1})$, $B(q^{-1})$ 和 $C(q^{-1})$ 是一步延迟算子 q^{-1} 的稳定的多项式,为方便计,不失一般性,设这些多项式的次数皆为 n, $\alpha(k)$ 是未知时变参数, $\overline{N}$ 是适当的常数, $e_{(k)}$ 是白噪声.

为估计未知参数 $\alpha(k)$ 的值,置

$$\phi(k)=[-y_{k-1},\ \cdots,\ -y_{k-n},\ u(k-h),\ \cdots,\ u(k-h-n)$$
$$\varepsilon(k-1),\ \cdots,\ \varepsilon(k-n)]^{\tau}$$
$$\varepsilon(k)=y_k-\phi(k)^{\tau}\theta-\hat{\alpha}(k)\overline{N}$$
$$\theta=[a_1,\ \cdots,\ a_n,\ b_0,\ b_1,\ \cdots,\ b_n,\ c_1,\ \cdots,\ c_n]$$

于是 $\hat{\alpha}(k)$ 满足

$$\hat{\alpha}(k)=\hat{\alpha}(k-1)+\frac{1}{\overline{N}}\{y_k-\phi(k)^{\tau}\theta-\overline{N}\hat{\alpha}(k-1)\}$$

$$=\frac{1}{\overline{N}}\{y_k-\phi(k)^{\tau}\theta\}$$

以下来寻求系统(8.13)的最小方差控制律:

由模型(8.13)得出

$$y_{k+h}=\frac{B(q^{-1})}{A(q^{-1})}u(k)+\frac{C(q^{-1})}{A(q^{-1})}e_{k+h}+\frac{\bar{N}}{A(q^{-1})}\alpha(k+m)$$

以及

$$e_k=\frac{A(q^{-1})}{C(q^{-1})}y_k-\frac{B(q^{-1})}{C(q^{-1})}q^{-m}u(k)-\frac{\bar{N}}{C(q^{-1})}\alpha(k)$$

把 $\frac{C(q^{-1})}{A(q^{-1})}$ 表示成如下的形式

$$\frac{C(q^{-1})}{A(q^{-1})}=D(q^{-1})+q^{-h}\frac{F(q^{-1})}{A(q^{-1})}$$

其中

$$D(q^{-1})=1+d_1q^{-1}+\cdots+d_{h-1}q^{-h+1}$$
$$F(q^{-1})=f_0+f_1q^{-1}+\cdots+f_{n-1}q^{-n+1}$$

经简单的计算不难得出

$$\begin{aligned}y_{k+h}=&\frac{B(q^{-1})D(q^{-1})}{C(q^{-1})}u(k)+\frac{F(q^{-1})}{C(q^{-1})}y_k\\&-\frac{F(q^{-1})\bar{N}}{A(q^{-1})C(q^{-1})}\alpha(k)+\frac{\bar{N}}{A(q^{-1})}\alpha(k+h)\\&+D(q^{-1})e_{k+h}\end{aligned}\tag{8.14}$$

显然 $D(q^{-1})e_{k+h}$ 与 $\frac{B(q^{-1})D(q^{-1})}{C(q^{-1})}u(k)+\frac{F(q^{-1})}{C(q^{-1})}y_k$ 是相互独立的. 而$-\frac{F(q^{-1})\bar{N}}{A(q^{-1})C(q^{-1})}\alpha(k)+\frac{\bar{N}}{A(q^{-1})}\alpha(k+h)$ 是由模型中的未知部分所形成的，我们假定它与系统的干扰也是独立的.如果 $y(k+h)$ 的希望值是 y_0 则由(8.14)式不难得出最小方差控制律为

$$\begin{aligned}u(k)=\frac{C(q^{-1})}{B(q^{-1})D(q^{-1})}\Big\{&y_0-\frac{F(q^{-1})}{C(q^{-1})}y_k+\frac{F(q^{-1})\bar{N}}{A(q^{-1})C(q^{-1})}\hat{\alpha}(k)\\&-\frac{\bar{N}}{A(q^{-1})}\hat{\alpha}(k+h|k)\Big\}\end{aligned}\tag{8.15}$$

其中 $\hat{\alpha}(k)$ 表示 $\alpha(k)$ 的估值，而 $\hat{\alpha}(k+h|k)$ 表示 $\alpha(k)$ 依据数据 $\{\hat{\alpha}(0),\hat{\alpha}(1),\cdots,\hat{\alpha}(k)\}$ 所得的向前 h 步的预报估值. 我

们可以应用多层递阶方法来寻求一系列 $\hat{a}(k+m|k)\ m=1, 2, \cdots, h$ 的估值[8]. 补偿自适应控制律的算法由公式(8.15)

$$\hat{a}(k)=\frac{1}{N}\{y_k-\phi(k)^\tau\theta\}$$

以及关于 $\{\hat{a}(k)\}$ 的向前 h 步的多层递阶预报算法所组成. 这种算法的优点是把参数自适应性由 $A(q^{-1})$, $B(q^{-1})$ 及 $C(q^{-1})$ 的参数转移到参数 $\alpha(k)$ 上. 这样既考虑了控制律的自适应性,又照顾了系统的稳定性.

§ 8.5 多层递阶参数自适应控制在造纸机上的应用

把前面所得到的自适应控制方法应用于"3150"新闻纸机造纸过程的控制,可以得到良好的效果.

表现新闻纸的质量的指标很多,但其中最重要的是定量(又称基重),即单位面积纸张的重量和水份,即纸张中的相对含水量.若能实现较好地控制这两个质量指标,定会收到较大的社会经济效益.

定量和水份是相互耦合的. 送料阀门和蒸汽管道阀门是调节定量和水份的两个重要的执行机构. 此外,纸浆的浓度、纸机一段烘缸的蒸气压力、纸机的车速等,对纸张的定量和水分都有较明显的影响. 我们用 $y_1(k)$ 表示纸张 k 时刻的定量, $y_2(k)$ 表示纸张在 k 时刻的水份; $u_1(k)$ 表示送料阀门的开度 (以电压表示), $u_2(k)$ 表示蒸气管道阀门的开度 (以电压表示); $s_1(k)$ 表示纸浆的浓度, $s_2(k)$ 表示纸机一段烘缸的蒸气压力, $s_3(k)$ 表示纸机的车速. 于是有

$$\mathbf{y}(k)=\begin{pmatrix}y_1(k)\\y_2(k)\end{pmatrix}\quad \mathbf{u}(k)=\begin{pmatrix}u_1(k)\\u_2(k)\end{pmatrix}\quad \mathbf{s}(k)=\begin{pmatrix}s_1(k)\\s_2(k)\\s_3(k)\end{pmatrix}$$

依据对造纸工艺过程的分析, 可以得到纸张定量水份控制系统的数学模型如下

$$\mathbf{y}(k)+A_1(k)\mathbf{y}(k-1)+A_2(k)\mathbf{y}(k-2)=B_1(k)\mathbf{u}(k-1)+B_2(k)\mathbf{u}(k-2)+C(k)\mathbf{s}(k-1) \tag{8.16}$$

此处

$$A_i(k)=\begin{pmatrix}a_{11}^i(k) & a_{12}^i(k)\\ a_{21}^i(k) & a_{22}^i(k)\end{pmatrix}\quad i=1,2$$

$$B_i(k)=\begin{pmatrix}b_{11}^i(k) & 0\\ 0 & b_{21}^i(k)\end{pmatrix}\quad i=1,2$$

$$C(k)=\begin{pmatrix}c_{11}(k) & c_{12}(k) & c_{13}(k)\\ c_{21}(k) & c_{22}(k) & c_{23}(k)\end{pmatrix}$$

显然能够把(8.16)式分离成两个广义单输出模型. 以定量模型为例,我们有

$$\begin{aligned}y_1(k)&+a_{11}^1(k)y_1(k-1)+a_{12}^1(k)y_2(k-1)+a_{11}^2y_1(k-2)\\&+a_{12}^2(k)y_2(k-2)=b_{11}^1(k)u_1(k-1)+b_{11}^2(k)u_1(k-2)\\&\qquad+c_{11}(k)s_1(k-1)\\&\qquad+c_{12}(k)s_2(k-1)\\&\qquad+c_{13}(k)s_3(k-1)\end{aligned} \tag{8.17}$$

首先确定参数 $a_{11}^1(k)=\tilde{a}_{11}^1$, $a_{11}^2(k)=\tilde{a}_{11}^2$, $b_{11}^1(k)=\tilde{b}_{11}^1$, $b_{11}^2(k)=\tilde{b}_{11}^2$, 使多项式 $1+\tilde{a}_{11}^1q^{-1}+\tilde{a}_{11}^2q^{-2}$ 和 $\tilde{b}_{11}^1+\tilde{b}_{11}^2q^{-1}$ 都是稳定的.

假定没有对纸浆的浓度进行在线检测, 取 $s_3(k-1)=s_3$ 为浓度的平均值.

置

$$z(k)=y_1(k)+\tilde{a}_{11}^1y_1(k-1)+\tilde{a}_{11}^2y_1(k-2)-\tilde{b}_{11}^2u_1(k-2)$$

$$\boldsymbol{\varphi}(k-1)=[-y_2(k-1),\ -y_2(k-2),\ s_1(k-1),\ s_2(k-1),\ s_3]^\tau$$

$$\boldsymbol{\theta}(k)=[a_{12}^1(k),\ a_{12}^2(k),\ c_{11}(k),\ c_{12}(k),\ c_{13}(k)]^\tau$$

于是模型(8.17)可以写成

$$z(k)-\tilde{b}_{11}^1u_1(k-1)=\boldsymbol{\varphi}(k-1)^\tau\boldsymbol{\theta}(k) \tag{8.18}$$

未知时变参数 $\theta(k)$ 的估值 $\hat{\theta}(k)$ 可以用递推算法求得，即

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{1}{\|\varphi(k)\|^2}\varphi(k)\{z(k)-\tilde{b}_{11}^1u_1(k-1)-\varphi(k)^\tau\hat{\theta}(k-1)\} \qquad (8.19)$$

为了寻求控制律，我们把模型(8.18)改写成

$$z(k+1)-\tilde{b}_{11}^1u_1(k)=\varphi(k)^\tau\theta(k+1)$$

对于由递推算法(8.19)得出的 $\hat{\theta}(k+1)$ 显然有

$$z(k+1)-\varphi(k)^\tau\hat{\theta}(k+1)=\tilde{b}_{11}^1u_1(k)$$

然而，在寻求控制律 $u_1(k)$ 时，$z(k+1)$ 是未知的．所以 $\hat{\theta}(k+1)$ 也是未知的．我们可以用 $\hat{\theta}(k+1)$ 的多层递阶预报估值 $\hat{\theta}(k+1|k)$ 来代替它．

如果 $y_1(k+1)$ 的希望值是 y_1^0，把与之相应的 $z(k+1)$ 记成 $z_0(k+1)$，于是可以得出纸张定量的自适应控制律是

$$u_1(k)=\frac{1}{\tilde{b}_{11}^1}\{z_0(k+1)-\varphi(k)^\tau\hat{\theta}(k+1|k)\}$$

同样可以导出水份的自适应控制律．应用这种方法，对造纸机进行控制，取得了令人满意的效果．

例如，取定量和水份的目标值分别为

$$y_1^0=50 \qquad y_2^0=7$$

控制的结果见表 8.1．

表 8.1

序　号	1	2	3	4	5	6	7
定量 (g/m²)	49.00	48.87	48.99	49.13	49.96	49.58	49.49
水份 (%)	6.78	6.81	6.7	6.81	7.51	7.2	7.1

序　号	8	9	10	11	12	13	14
定量 (g/m²)	50.05	49.83	50.61	50.55	50.71	50.84	50.41
水份 (%)	7.06	7.0	7.45	7.31	7.14	7.45	7.07

序　号	15	16	17	18	19	20
定量 (g/m²)	50.72	50.01	50.12	50.58	50.29	50.50
水份 (%)	7.03	6.61	6.78	7.03	7.01	7.01

参 考 文 献

[1] J. L. Doob, Stochastic Processes, Wiley, New York, 1953.
[2] K. J. Åström, Introduction to Stochastic Control Theory, Academic Press, 1970.
[3] B. Wittenmark, A self-tuning predictor, *IEEE Trans.*, AC-19, No. 6, 1974.
[4] M. M. Loève, Probability, Theory, 3rd ed., Van Nostrand, New York, 1963.
[5] L. Ljung, Consistency of Least Squares Identification Method, *IEEE Trans.*, AC-21, No. 5, 1976.
[6] G. C. Goodwin and R. L. Payne, Dynamic System Identification, Experiment Design and Data Analysis, Academic Press, 1977.
[7] E. J. Hannan, Recursive Estimation Based on ARMA Models, *The Ann. of Statis.*, Vol 8, No. 4, 1980.
[8] L. Ljung, Analysis of Recursive Stochastic Algorithms, *IEEE Trans.* AC-22, No. 4, 1977.
[9] H. J. Kushner and D. S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Springer-Verlag, New York, 1978.
[10] T. Söderström, L. Ljung and I. Gustavssion, A Theoritical Analysis of Recursive Identification Methods, *Automatica*, Vol. 14, 1978.
[11] H. Robbins and S. Monro, A Stochastic Approximation Method, *Ann. Math. Stat.*, Vol 22, No. 1, 1951.
[12] J. Kiefer and J. Wolfowitz, Stochastic Estimation of the Maximum of a Regression Function, *Ann. Math. Stat.*, Vol 23, No. 3, 1952.
[13] A. Dvoretsky, On Stochastic Approximation, Proc. 3rd Berkeley Symp., *on Math. Statist. and Probab.*, Vol 1, 1956.
[14] M. B. Priestley, Spectral Analysis and Time Series, Vol 2, Academic Press, 1981.
[15] H. Tong, On a Threshold Model in Pattern Recognition and Signal Processing, Sijthoff and Noordhoff, Amsterdam, 1978.
[16] C. W. J. Granger and A. P. Andersen., An Introduction to Bilinear Time Series Models, Vandenhoek and Ruprecht, Göttingen, 1978.
[17] M. B. Priestley, State-dependent models: a General Approach to Non-linear Time Series Analysis, *Time Series Analysis*, Vol. 1, No. 1, 1980.